Undergraduate Research Semester Internship Fall 1999 Abstracts and Research Papers

S. Brown, L. Capaldi, M.A. Garcia, S.S. Kahn, D.N. Monbleau, D. Nault, N. Nicks, M. Ho, N.L. Tran

December 1, 1999

U.S. Department of Energy

Lawrence
Livermore
National
Laboratory

DISCLAIMER

This report has been reproduced directly from the best available copy.

Available electronically at http://www.doe.gov/bridge

Available for a processing fee to U.S. Department of Energy
and its contractors in paper from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-mail: reports@adonis.osti.gov

Available for the sale to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/ordering.htm

OR

Lawrence Livermore National Laboratory
Technical Information Department's Digital Library
http://www.llnl.gov/tid/Library.html

Foreword

The Undergraduate Research Semester (URS) program provides a unique and challenging off-campus research opportunity for upper-division university undergraduate and pre-grad-school students in science, mathematics, and engineering. This internship program is a partnership between Lawrence Livermore National Laboratory, Sandia National Laboratories (New Mexico and California), and Los Alamos National Laboratory, to provide 75+ (annually) science and engineering undergraduates a rich research experience in this collaborative program.

The URS project supports the DP mission through ensuring a scientifically and technically literate citizenry, and contributes to the development of a highly skilled, diverse scientific workforce, with experience, exposure and increased awareness and support for the DP science/technology and Science Based Stockpile Stewardship.

U R
S ★
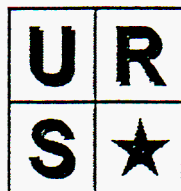
Science & Technology Education Program
Fall 1999 URS Internship Program
at
Lawrence Livermore National Laboratory

Authors:

- Steven Brown
- Laura Capaldi
- Miguel A. Garcia
- Sandia S. Kahn
- Davienne Noelle Monbleau
- David Nault
- Nathaniel Nicks & Mai Ho
- Ngoc Lien Tran

Affiliations:

Steven Brown, University of California, Berkeley; Laura Capaldi, Brown University; Miguel A. Garcia, University of California, Davis; Sadia S. Kahn, California State University, Fullerton; Davienne Noelle Monbleau, Rensselaer Polytechnic Institute; David Nault, University of Cincinnati; Nathaniel Nicks, University of Washington; Mai Ho, Georgia Institute of Technology; Ngoc Lien Tran, Contra Costa Community College

Undergraduate

```
┌───┬───┐
│ U │ R │
├───┼───┤
│ S │ ★ │
└───┴───┘
```

Research Semester

Table of Contents

Characterization of the Mouse UvrD Gene Encoding a Putative Helicase

Laura Capaldi
Brown University,
Biology

ABSTRACT

We previously identified the mouse UvrD gene based on a motif sequence common to helicase genes in bacteria and yeast. The object of our current study is to characterize this mammalian gene and analyze the structure and function of its protein product. After determining the complete sequence of a cDNA for the mouse UvrD gene, we prepared the cDNA for protein expression in bacteria. Once the protein is purified we will be able to investigate its biochemical function. We predict that the protein will possess helicase (DNA unwinding) activity. Knowledge concerning mammalian UvrD structure and function will aid in our understanding of its function in human health and disease. Further investigation may lead to the discovery that certain mutations in the UvrD gene result in faulty DNA repair, replication, or recombination. Studies may also link a defective UvrD gene with a disease, such as cancer. Once a defective gene is linked with a disease, researchers will be able to screen humans with mutations in the homologous human UvrD gene and thus monitor those having a genetic predisposition for the disease.

Parallel Object-Oriented File I/O
For the Sapphire Project

David Nault
University of Cincinnati
Computer Science

ABSTRACT

The purpose of the Sapphire Project is to provide researchers with a set of tools for interactive exploration of large, complex, multidimensional data sets. As an initial test-bed application we are working with astronomical data from the VLA FIRST Survey in an effort to classify radio emitting galaxies with a bent-double morphology. In order to solve this problem we are applying and extending ideas from data mining and pattern recognition in our development of a scalable, parallel application toolkit in C++. Various parts of the application toolkit will be discussed, with the primary focus on the file input/output system and the parallel communication infrastructure.

Davienne Noelle Monbleau
Rensselaer Polytechnic Institute
B.S. Physics, May 1999

ABSTRACT

Our ultimate objective is to develop a three-dimensional stellar evolution code that will emulate the thermodynamic configuration of non-spherical stellar systems and their evolution through time. In preparation, we are testing the accuracy of the basic input physics of our one-dimensional code by comparing sound-speed virus depth from an evolved solar model to observed helioseismic data. The code may then be applied to other stars to provide more precise evolutionary models and further the development of a three dimensional code.

Timing Circuit For Piston Engine Research

Steven Brown
University of California, Berkeley
Engineering Physics

ABSTRACT

LLNL and UC Berkeley are exploring power generation from a piston engine fueled by natural gas augmented with, for example, di methyl ether (DME). The DME is added to help start the ignition of the natural gas. However, the time of the combustion event is remarkably sensitive to the amount of additive DME. Small changes in the DME mole fraction lead to large changes in timing. The timing of this combustion event, relative to top-dead center (TDC) is erratic. As we wish to capture the combustion event with data acquisition computers and oscilloscopes, we need a timing system that allows great flexibility in defining when the data acquisition should commence. We have developed two approaches: The analog system is based on phototransistors that are illuminated by light reflected from a rotating shaft that is illuminated by light emitting diodes (LED). The shaft is covered with black tape, except for an exposed shiny metal strip. As light reflects from the exposed shiny metal, a timing pulse is generated. Adjustment of the timing is done by rotating the LED-detector combination to a new observation position. This rugged analog system has worked well. The digital system counts pulses sent from a 'shaft encoder' that starts pulses at BDC. The digital system has worked on a test bed, but is yet to be attached to the single cylinder engine, in part due to the success of the analog system. We will give examples of high quality data collected with system 1.

Removing Nitrate and Perchlorate From Groundwater With Anaerobic Microbial Films

Nathaniel Nicks
University of Washington
B.S. Chemical Engineering, June 1999
&
Mai Ho
Georgia Institute of Technology
B.S. Chemical Engineering, June 1999

ABSTRACT

The purpose of this project was to design a remediation unit that will lower nitrate and perchlorate concentrations in the groundwater at the DOE test facility Site 300. The design process included evaluation of an existing bioreactor, and consulting with LLNL engineers, researchers, and technicians to design a more versatile bioreactor. The bioreactor will be a secondary treatment unit following a conventional solvent removal unit. The design consists of two packed towers connected in series. Microorganisms are added to the bioreactor to colonize and form a biofilm on the packing material. The addition of ethanol to the reactor will allow the biofilm to metabolize the nitrate and perchlorate contaminants into harmless components: nitrogen gas, chloride ion, and water.

The Prediction of Biochemical Acid Dissociation Constants Using First Principles Quantum Chemical Simulations

Ngoc Lien Tran
Contra Costa Community College
Chemistry

ABSTRACT

The motion of protons (H^+) between molecules is a vital part of many chemical processes. This process of proton transfer is defined by a chemical's acid dissociation constant (pK_a). The goal of this study is to evaluate existing quantum chemical methods to predict pK_a. We used gradient-corrected density functional theory with a B3LYP functional with different basis sets and solvent models to determine the trade-off between the methods' accuracy and computational cost. Our results show a good agreement between experimental and calculated dissociation constants with R^2 up to 0.997 for carboxylates and 1.000 for phosphates, indicating the feasibility of accurately predicting acid dissociation constants using quantum chemical methods.

Novel Diode-Pumped Solid State High Average Power Laser Design

Miguel A. Garcia
University of California, Davis
B.S. Physics, June 1999

ABSTRACT

A proprietary laser design uses diode pump-light introduced at the edges of the device. Using ray trace models, we optimized the optical geometry and dimensions of the device to produce the highest efficiency and power output.

Crystallization of Bacterial Toxins For X-ray Crystallographic Analysis

Sadia S. Kahn
California State University, Fullerton
B.S. Biochemistry, August 1999

ABSTRACT

The purpose of this study is to find suitable crystallization conditions for two bacterial toxins, mutant *Staphylococcus aureus* enterotoxin B (SEB) and *Clostridium* botulinum neurotoxin A light chain (BoNT A). Our anticipated goal is to produce crystals amenable to X-ray diffraction analysis. Initial crystallization conditions were found through random screens and refined in further experiments to improve crystal quality.

Initial results indicate that:
- For SEB:
    - Ethanol crystallized the protein at low and high concentrations as well as at low and high pH. Co-precipitants will be tested in future experiments to improve crystal quality.
    - Mg^{2+} or Cd^{2+} in combination with PEG 10K yielded small crystals. Both metal and PEG 10K concentrations will be varied in future experiments to improve crystal quality and size. In addition, Mg^{2+} in combination with HEPES buffer (pH 8.5) yielded crystals. Various buffers at different pH will also be tested.

- For BoNT A:
    - A specific combination of PEG MME 5K and Na succinate buffer (pH 4.5) yielded needle-shaped crystals. Experiments varying PEG MME 5K concentration and buffer pH are in progress to improve crystal quality.
    - A specific combination of $(NH_4)_2SO_4$, Na acetate buffer (pH 6.5), and glycerol yielded spherical crystals. The concentrations of the reagents and buffer pH will be varied in further experiments.

Timing Circuit for Piston Engine Research[*]

Steven A. Brown

University of California, Berkeley

Lawrence Livermore National Laboratory
Livermore, California 94551

12/14/1999

Timing Circuit for Piston Engine Research

Steven A. Brown
University of California, Berkeley
Mechanical Engineering

ABSTRACT

LLNL and UC Berkeley are exploring power generation from a piston engine fueled by natural gas augmented with, for example, di-methyl-ether (DME). The DME is added to help start the ignition of the natural gas. However, the time of the combustion event is remarkably sensitive to inlet temperature, pressure, equivalence ratio, and engine speed to the amount of additive DME. Small changes in the DME mole fraction lead to large changes in timing. The timing of this combustion event, relative to top-dead center (TDC) is erratic. As we wish to capture the combustion event with data acquisition computers and oscilloscopes, we need a timing system that allows great flexibility in defining when the data acquisition should commence. We have developed two approaches:

- The analog system is based on phototransistors that are illuminated by light reflected from a rotating shaft that is illuminated by light emitting diodes (LED).
- The digital system counts pulses sent from a 'shaft encoder' that starts pulses at bottom-dead center (BDC).

We will give examples of high quality data collected with the analog system.

Introduction

On highway and off highway, heavy-duty diesel engines contribute a significant proportion of the total NO_X and particulate emissions produced in the most Urban areas. This includes emissions from trucks, earthmoving equipment, locomotives and marine applications. The primary attractions of diesel powered equipment are high fuel efficiency, low first cost, and engine durability. Unfortunately, these advantages are offset by the inherently high emissions levels produced by nonpremixed combustion. There are a number of low emissions alternatives to the Diesel engine including fuel cells, gas turbines, and spark-ignited engines with exhaust after treatment. None of these power sources have made significant inroads towards displacing Diesel engines due to their shortcomings in one of the above categories. In the near term, none of the Diesel alternatives so far proposed will achieve substantial market penetration.

Homogeneous Charge Compression Ignition (HCCI) engines have been under development for the last 20 years. HCCI describes a combustion process in which a well-mixed charge of fuel and air is brought to ignition by compression in a traditional reciprocating engine. These processes combine the high efficiency of the Diesel cycle with very low NO_X and essentially zero particulate emissions. HCCI engines achieve this performance by utilizing the best characteristics of Diesel and spark ignition engine operation. See Table 1. The HCCI engine is similar to a Diesel engine in that it operates at a high compression ratio without intake throttling. These characteristics produce fuel economy on par with modern Diesel engines. Similar to spark ignited engines, though, the HCCI engine does not rely on diffusion burning of the fuel, but rather is based on a uniform combustion process. These properties of HCCI engines remove the high temperature, stoichiometric and rich combustion regions, which generate the high emissions levels in Diesel combustion. As opposed to other Diesel alternative technologies, this concept does not require the development of new hardware or exotic materials, but is simply a modification of current production engines. For this reason, it will be competitive with current market leaders in terms of engine cost and durability, as well as fuel economy.

Advantages of HCCI

	Diesel	Spark	HCCI
Efficiency	45 %	28 %	45 %
Emissions			
NO_x	300ppm	50ppm**	15ppm
Particulate	High	~Zero	~Zero
CO	Low	Medium *	High *
Hydro-Carbons	Low	Medium *	High *
Types of Fuels			
Natural Gas	No	Yes	Yes
Gasoline	~No	Yes	Yes
Diesel Fuel	Yes	No	Maybe

* Can be Remove with Oxidation Catalyst   ** With use of 3-way Catalyst

Table 1

To this point, limited application of the HCCI principle has been made in the commercial domain due to the two historical drawbacks of the technology; difficulty in engine control and limited load/speed capability. With recent developments in sensors and controls, these historical problems are now being solved. Work at a number of research institutions around the world has verified that HCCI engines can be run with very low NOx and particulate emissions while producing thermal efficiency equal to or greater than current Diesel engines (we have at least 100 references spanning the last 5 years.).

Explanation of Methods Used

The difficulties in engine control arises from the fact that there is no spark plug, no fuel injector, and thus no direct control of the ignition timing in an HCCI engine. Ignition occurs when the trapped mixture reaches a critical conditions, which is a function of inlet temperature, pressure, equivalence ratio, and engine speed. In order to insure optimum engine performance all of these parameters must be carefully controlled to produce the desired engine speed and load. The use of novel sensors and feedback mechanisms under development by the author, and the HCCI team, will allow the application of the HCCI process over a range of operating conditions.

In particular we need to know how close to top dead center (TDC) is the HCCI explosion, seen as the sudden change in the pressure near TDC. As seen in Fig. 1.

The analog system is based on phototransistors that are illuminated by light reflected from a rotating shaft that is illuminated by light emitting diodes (LED). The shaft is covered with black tape, except for an exposed shiny metal strip. As light reflects from the exposed shiny metal, a timing pulse is generated. Adjustment of the timing is done by rotating the LED-detector to a new observation position.
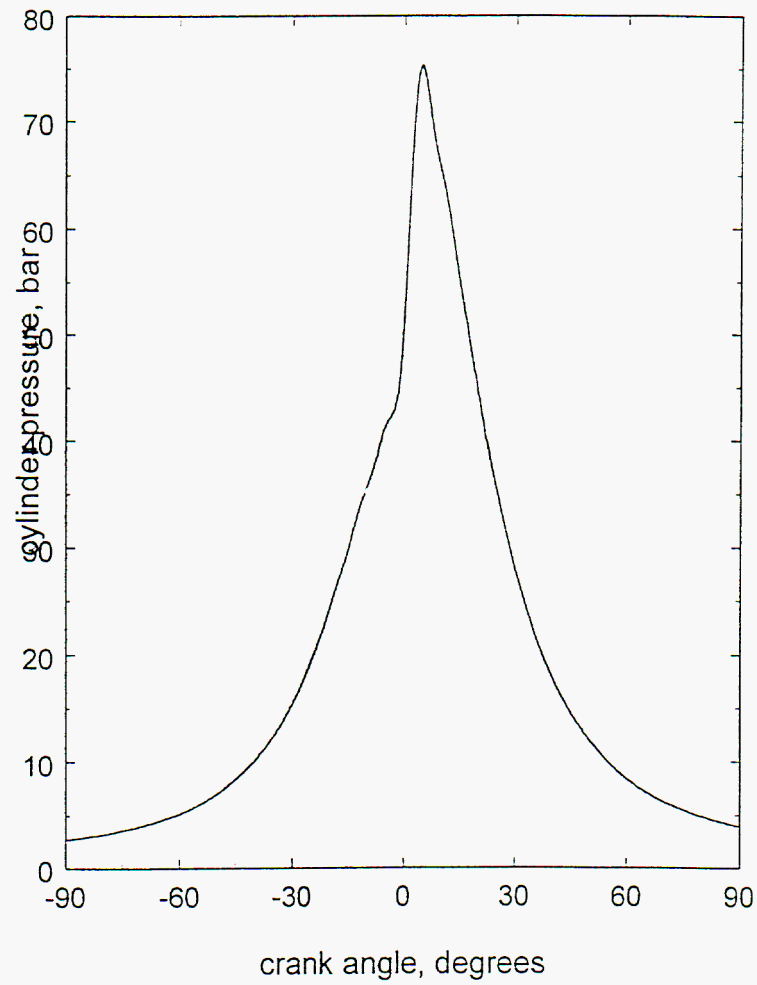
Figure 1

Figure 1 is a pressure verse time history for an HCCI combustion event in a CFR engine with compression ratio of 17. TDC is at zero degrees. The HCCI explosion is indicated by the sudden change in derivative at TDC (p ~ 44Bar).

Presentation of Data

The analog timing circuit was attached to the Cooperative Fuels Research (CFR) Engine. This one cylinder HCCI engine was our test bed for much of the HCCI research done. Figure 2 shows an example of how the analog timing circuit allowed us to find crank angle of peak heat release for various fuel combinations and air temperatures. Data such as in Fig. 2 is a prerequisite to control of the HCCI engine.
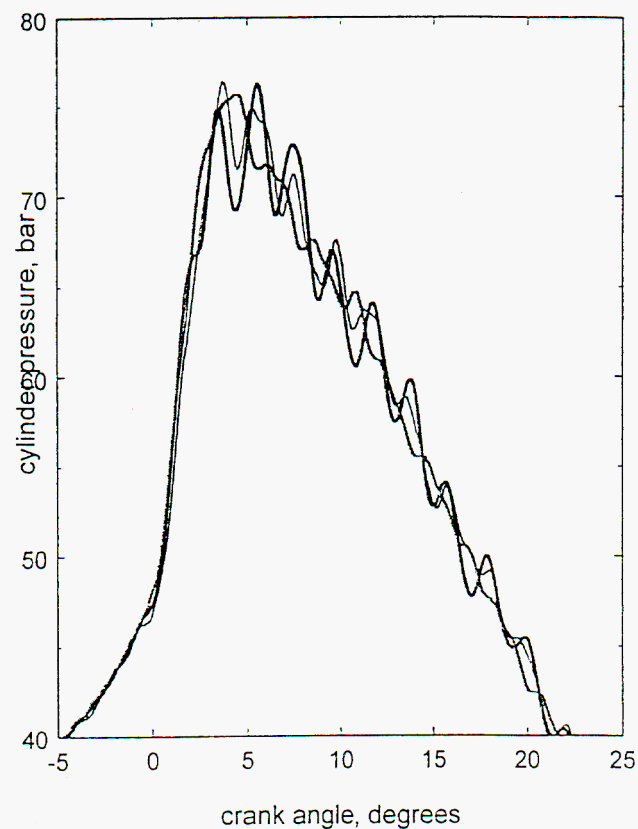
Figure 2

Several instantaneous pressure verse time events in an HCCI engine. Note that this Figure is an 'close up' of Figure 1 and thus one can see better the change in derivative associated with the HCCI combustion event.

Discussion and Analysis of Results

This rugged analog system has worked well. With the use of better patterns, this system has preformed to satisfaction. We were able to use it for all the data acquisition that we needed to do on the CFR engine

The analog system has its limitations. The future of HCCI will involve computers actively controlling many factors in the combustion process to insure that start of combustion (SOC) occurs as close to TDC as possible. This would require a digital system to give the controlling system the information about SOC fast enough so that they computer could correct for any deviation from TDC. The digital system counts pulses sent from a 'shaft encoder' which would start sending pulses at BDC. Fig. 3 is a schematic of what one possible digital circuit would look like. However it has not been pursued, in part due to the success of the analog system.
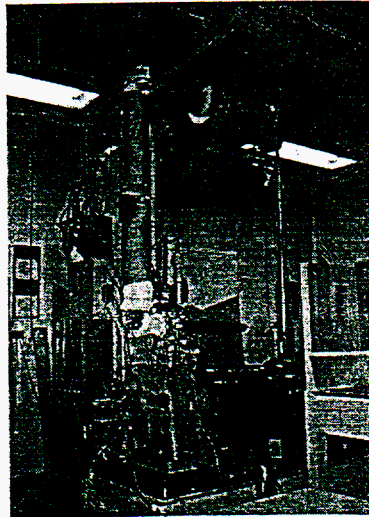
CFR Engine

Figure 3

Conclusion

There is no doubt that HCCI engines have a wonderful future in the 21st century. The question with these engines is timing. We have successfully used an analog timing circuit to run a one cylinder HCCI engine. For research purposes we have found an answer to timing.

However, for future HCCI engines, ones that could be used commercially, the answer is digital. As we can see in Figure 4, just a few crank angle degrees can make a huge difference. This shows that damaging peak pressures can be avoided when the combustion event is delayed relative to TDC. At some point we must develop a working digital time to be used in conjunction with a controlling computer.
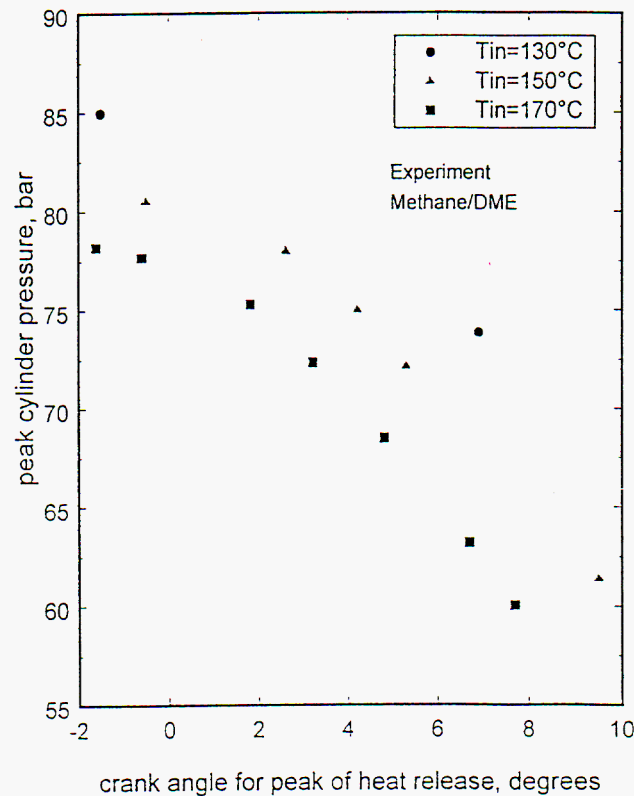
Figure 4

Characterization of the Mouse *UvrD* Gene Encoding a Putative Helicase *

Laura Capaldi, Brown University
Michael Thelen, Mark Shannon, Jim George
Lawrence Livermore National Laboratory
Livermore, California 94551

December 10, 1999

Abstract

We previously identified the mouse *UvrD* gene based on a motif sequence common to helicase genes in bacteria and yeast. The object of our current study is to characterize this mammalian gene and analyze the structure and function of its protein product. After determining the complete sequence of a cDNA for the mouse *UvrD* gene, we prepared the cDNA for protein expression in bacteria. Once the protein is purified we will be able to investigate its biochemical function. We predict that the protein will possess helicase (DNA unwinding) activity. Knowledge concerning mammalian UvrD structure and function will aid in our understanding of its function in human health and disease.

Introduction

DNA helicases are enzymes that use energy derived from ATP to unwind DNA strands during DNA replication, repair, and recombination (see figure 1).
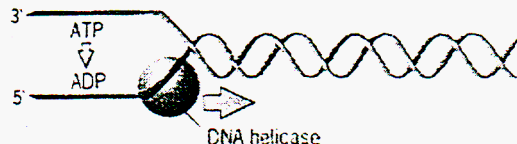
Figure 1 DNA helicase unwinding DNA. *(taken from Snustad, Simmons, and Jenkins, 1997)*

Since these metabolic processes occur in organisms ranging from bacteria to humans, it is not surprising that DNA helicase sequence motifs have been conserved among these organisms. Helicase sequence motifs are specific amino acid sequences that are present in most helicases (see figure 2).


```
                          Motif I                 Motif Ia
Mouse UvrD →  mUvrD    1  —GTGKTSTLVKYAEKWSQ————SRFLYVTFNKSIAKQAELVFPSNVICKTFHS——MA
Yeast UvrD →  yUvrD    1  —GTGKTKALLEFAKSRPK————DKILYVAENKAAKEDAELRFPFNVKCSTMHG——LA
 erial UvrD → bUvrD    1  GAGTGKTTIITNRFAYLVNKEKVDPSRILAITFTKKAAKEMQFRILKLIDSSLAEKTNIY

                          Motif II                  Motif III
              mUvrD  157  MTHDGYLKLWQLSKPLLASFDAIFVDEAQDCTPAIMNIVLSQPCG——KIFVGDPHQQIY
              yUvrD  164  DAYMKLLHLYEFPN—IFSKYDYILFDEAQDFTPCMVDLIYRQKHAR——IVIVGDAHQCIY
              bUvrD  174  KTCNLLSNDNDLLNQWSEQFQHILVDEFQDTNQIQYELIKMLVTKNKNLFLVGDNNQMIY

                          Motif IV
              mUvrD  214  TFRGAVNALF————TVPHTHVFYLTQSFRFGVEIAYVGATILDVCKRVRFKTLVGGN
              yUvrD  221  GFRGANACAFNEN——LYPSTKQLCLTKSFRFGNSVAKYANFLISLKGENVKLKGVQND
              bUvrD  234  RWRGAVHGIITALKHDFNVPKSNEFFINQNYRCDQNILAVANQILLKIMAYEKQVKTEKN

                          Motif V
              mUvrD  382  IPELVERIER————————CHIDDLDFAEY——ILGTVHKAKGLEFDTVHVLDD
              yUvrD  391  SESFISRLDN————————CEKRLMDSKDDGDNGIILATAHQSKGLEWDNVQLGND
              bUvrD  474  LWKKLTEFQKDKTEFSLSEFITSLALEFDSIIENSSDTINLLTVHAAHGLFFFAVFIVGM

                          Motif VI
              mUvrD  424  FVKVPCARHNLAQLPHFRVESFSEDEWNILYVAVTRAKKRLIMTKSLENILTLAG——EYF
              yUvrD  438  FR————PKFDSVSFSRIGSSRYLQEEINILYVALTRAKKRLILNDTITKLYALECGLVRF
              bUvrD  534  NQ————GDFPLFISQNQNDEQHLIDELKIFYVAITRAKRFLFITAVLQINNNSIKP-SSF
```

Figure 2 Conserved helicase motifs found in a mouse, yeast, and bacterial protein.

Our lab has identified a mouse gene, now called *UvrD*, with motifs conserved in bacterial, yeast, and human helicase genes. This gene was found through a database search using the sequence for bacterial *UvrD*. The yeast gene, *YAY5*, was obtained and was then used to search for other genes with homologous sequences. The search resulted in a 700 base pair mouse EST (Expressed Sequence Tag). It is this gene that the term

"mouse *UvrD*" refers to (see figure 3). The object of this project is to determine its sequence and prepare it for protein expression. We hypothesize that the mouse UvrD protein is a helicase based on its sequence homology to other helicases.

Bacterial *UvrD*

↓

Yeast Gene (*YAY5*)

↓

700 base pair mouse EST (mouse *UvrD*)
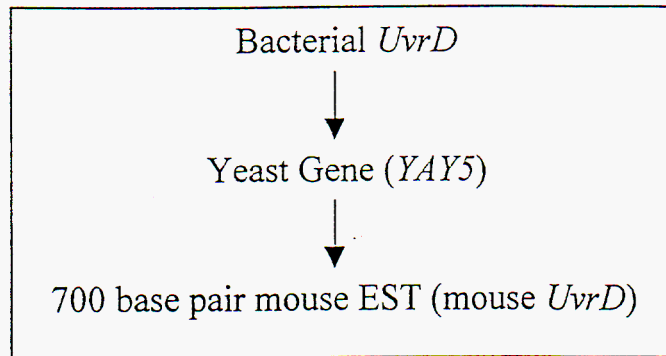
Figure 3 Steps taken to find the putative mouse helicase using a database search.

Methods

Determining the Length of the Mouse *UvrD* Transcript

The length of the mouse *UvrD* transcript is determined by performing an mRNA blot. In an mRNA blot, the 700 base pair mouse EST is allowed to hybridize to the corresponding mRNA on the gel. A size estimate is then established based on comparison to a standard in the gel.

Screening a cDNA Library

Colony hybridization is the method employed to screen a cDNA library. Using this method, a radioactive cDNA is used as a hybridization probe for a testis cDNA library.

RACE

RACE is the Rapid Amplification of cDNA Ends. It is one of the methods employed to obtain the full-length cDNA.

Steps to Sequencing

TA cloning

After isolating the target DNA from a gel, we subject it to TA cloning in which the DNA of interest is inserted into the PCR 2.1 Vector using ligase.

Transformation

The PCR 2.1 Vector along with its DNA insert is put into DH10 B bacterial cells using electroporation.

The cells are then grown on plates containing X-Gal and ampicillin. This method provides an easy way to identify colonies that have the vector with the desired DNA insert. Only the intact vector (missing the DNA insert) has the gene to covert X-Gal into B-Galactosidase, which causes the colony to take on a blue appearance. The DNA insert, however, breaks the vector at that gene; thus, white colonies have the vector with the insert and blue colonies do not (see figure 4). Also, only cells that have taken up the vector can grow on the plate because the vector has an ampicillin resistance gene.
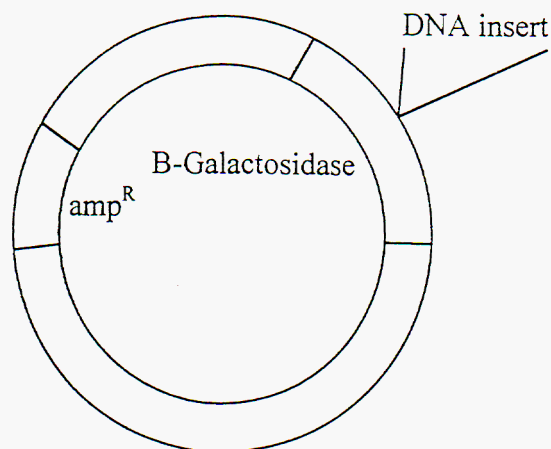
Figure 4 PCR 2.1 Vector

The next day, white colonies are selected and grown overnight in LB with ampicillin.

Mini-Prep

Using the reactants from the Wizard Kit, the vector with insert can be isolated from the bacteria. A restriction digest is performed and the products are separated by electrophoresis to be sure that the vector contains the insert.

Standard Sequencing Reaction

3 μl of the isolated vector with insert
1.5 μl of a primer (primers are designed to portions of the gene whose sequences are already known)
1.5 μl water
4 μl reaction mix (contains DNA polymerase, deoxynucleotides, and fluorescently tagged dideoxynucleotides)

The above cocktail is placed in a PCR machine overnight. The DNA of interest is then precipitated after addition of ethanol and loaded onto a polyacrylamide gel where the products are separated based on size by gel electrophoresis.

The DNA is then sequenced based on a variation of the chain terminator method, known as automated DNA sequencing, in which fluorescently tagged dideoxynucleotides are detected by a computer.

Preparing the cDNA for Expression

The cDNA for mouse *UvrD* and the pET 30a vector are cut with the restriction enzymes *Bam* H1 and *Not* 1. The insert and the vector are then ligated together (see figure 5). The new intact vector is now ready for expression in bacteria.

The cells are grown on Kanamycin plates to assure that they have taken up the vector and a restriction digest is performed on the vector to guarantee that the DNA of interest has been inserted into it. The DNA of interest is also resequenced to make certain that no mutations have arisen in the sequence.
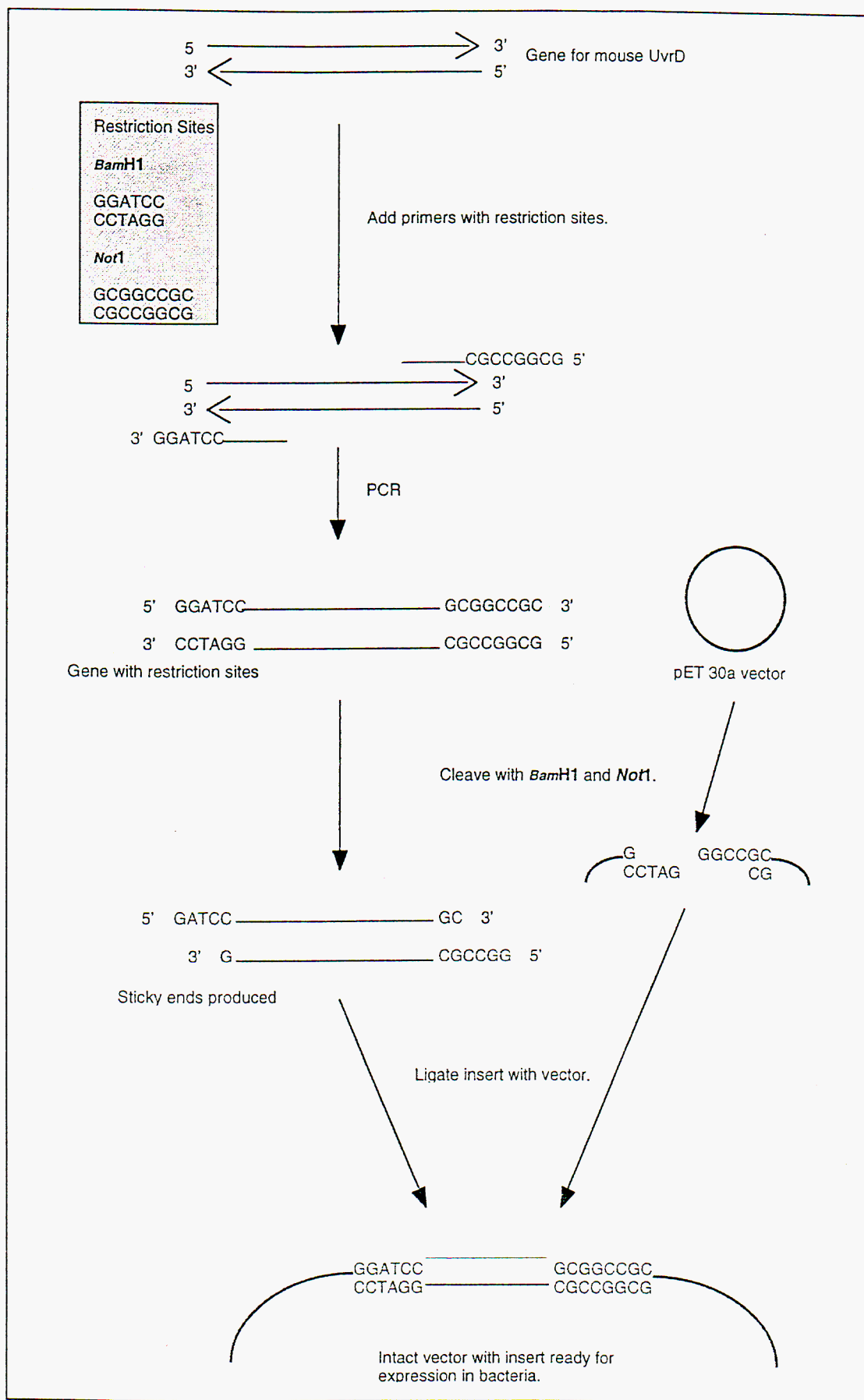
Figure 5 Incorporation of the cDNA into an expression vector.

Extracting Protein from the Cells

Epicurian Coli BL21-Codon Plus-RIL Competent Cells are the cells used to express the protein of interest. The transformation reaction is heat pulsed for 20 seconds. The cells are then grown to an optical density of 0.600, at which point bacterial expression of the target protein is induced by the addition of IPTG. After extracting the protein from the cells the soluble and insoluble proteins are separated and analyzed with a Western blot.

Affinity Chromatography is used to purify the protein, which has a His tag, over a Nickel column.

Results

The Complete Sequence

The mRNA blot indicated that the mouse UvrD transcript should be about 3.5 kb long (see figure 6).

The full length cDNA was obtained by screening a testis cDNA library and by performing RACE (see figure 7). The complete sequence of the cDNA for the mouse *UvrD* gene was then determined by automated DNA sequencing (see figure 8).
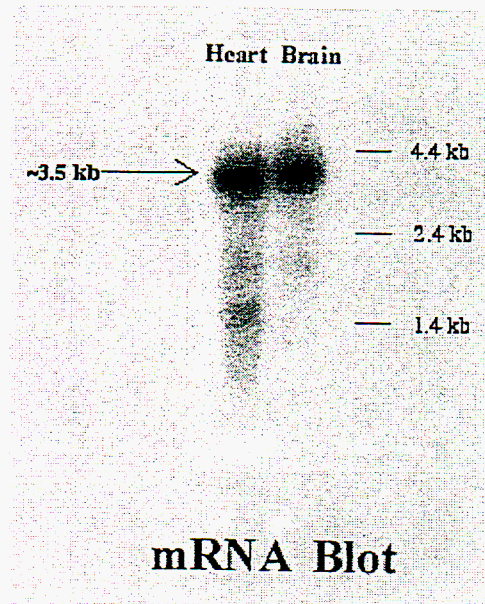
Figure 6 mRNa blot using RNA from heart and brain tissues in mice; size estimate based on comparison to standard.
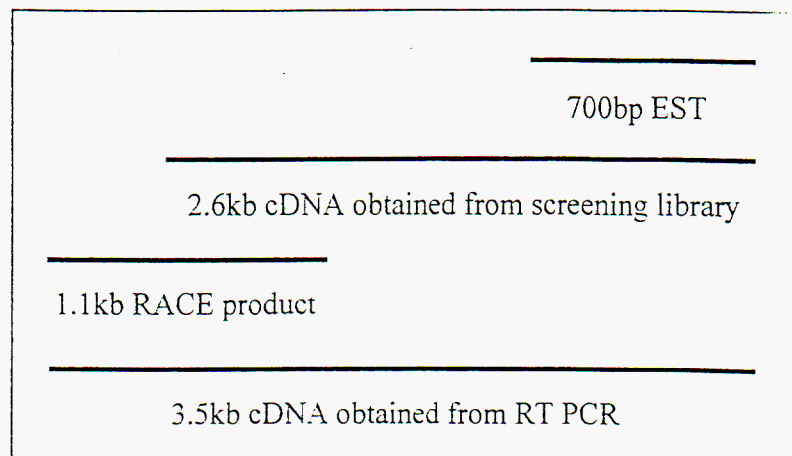
Figure 7 Steps to obtaining full length cDNA

```
CGCTGTCTGGGCCGAAGGAGGGGCTTGCTGGTCGGGGAAAGGCTGAATGTCCGGCGGCATCGCCTGGGTGGTGTCGCCGC
CGGGGTCCTGGGCCGGGCCGCCATAGTGCGTGAGACCGTGGACTTGGCGTGGGTGAGACGGTTTAAGCGGAAGCATCTTA
CAGTGGTCGACTGCCATCATTTGGCCCGGAGTCATTGGCTGTGACTCAGCCCTTCAGTCAGAGATGGACAAACAGAGACC
CTAACCATGGTCTCTATCCTAGACCCAGAACAAAAGGACGGAATAGGGGTCGAGGATGTCAGAGATATATTTCTGAGTTC
TTCCTAGCTGGCCACCAGCACTGTACCAATGACATGGCCAAAAGCAATTCTGTTGGCCAGGACAGCTGTCAGGACGCTGA
GGGTGACATGATTTTAACTGCAGAGAGCAGTTGTACTCTGCCTCAGGTGGATAATGGAGAAGCGAGGCTGGGCTCATCAG
GATCTGCACAGCCTGCCAGGAAGCGGGCTCACTGCTTTGAGGAAGCCACAGAGAGCGGCCAGTGGGATGGCGTTACTAAG
AAGACACCGCGCCATCGTTTGTTCCCATCGTGTTCAAGGCTTAGGGAGGCCAGGCAAGGGGCAGAGGACAGTTTGTCCCA
GTGCTCCCCAGTGCCGGGGGAAGCTGGCCGAGACATTGAGGACATTGGTCCTGATCCCTTGCCTGACTCATACTATGGGC
TTCTTGGAATGTTGCCTTGTCAGGAAGTGCCCAGCCACATTTGCAGGCTGCCAAGTGAAGTGCTGAGACACATCTTTGCT
TTCCTGCCAGTGGAGGACCTCTACTGGAATCTCAGCCTGGTGTGCCACCTGTGGAGGGAAATAATCAATGACCCGCTGTT
CATTCCTTGGAAGAAGCTATATCATCGATACTTGATCAATGAAGAGCAGGCTGTCAGCAAAGTGGATGGCATTCTCTCGA
GCCATGGCATAGAAAAGGATTCGGACCTATGTGTGCTGAACCTCATACGATACACGGCCACCACCAAATGCTCCCCAAGT
GTTGACCCTGAGAGGGTGCTGTGGAGCCTGAGGGATCATCCTCTACTTTTGGAGGCTGAGGCGTGCATGCGGCAACAACT
ACCTGACCTCTATGCAGCTGCTGGGGGCATTAACGTCTGGGCCTTGGTGGCAGCCATGGTGCTCCTCTCCAGCTGTGTGA
ACGACATCCAGCATCTGCTCTTCTGCCTCAGGAGACCCAGCTCCACAGTGACCATGCCAGATGTCACCGAGACCTTATAC
TGCATAGCTGTGCTCCTTTATGCCATGAGGGAGAAGGGGATTAACATCAGCAATAGGATTCACTACAACATTTTCTATTG
CCTATATCTTCAGGAGAATTCCTGTACTCAGGCCACCAAAGTTACAGAGGAGCCATCTGTCTGGCCAGGCAAGAAAACAT
CTATCCAGCTAACACATGAACAACAGCTGATTCTGAACCATAAGATGGAGCCTCTCCAGGTGGTAAAAATTATGGCATTT
GCAGGCACCGGGAAGACTTCTACCCTGGTCAAGTATGCTGAGAAGTGGTCTCAGAGCAGGTTTCTGTATGTCACGTTCAA
CAAGAGCATCGCAAAGCAGGCAGAGCTTGTCTTCCCCAGCAACGTCATCTGCAAGACCTTCCATTCCATGGCCTACAGTC
ACGTCGGGCGGAAGTACCAGCTGAAGAAGAAGTTGAATCTCTTCAAGTTGACACCCTTTATGGTTAACTCCGTCCTTGCT
GAAGGAAAAGGTGGATTTATAAGGGCCAAGTTAGTGTGTAAGACTTTAGAAAACTTCTTTGCATCAGCTGATGAAGAGCT
AACTATCGATCATGTGCCCATTTGGTGCAAGAATAGCCACGGGCAGAGAGTCATGGTGGAGCAGAGTGAGAAACTGAATG
GTGTCCTTGAAGCAAGCCGACTCTGGGACAACATGCGGAAGCTGGGGGAATGCAAAGAAGAGGCATACCAAATGACACAT
GATGGCTATTTGAAACTCTGGCAGCTGAGCAAACCTTTGCTTGCCTCTTTTGACGCCATCTTTGTGGATGAGGCCCAGGA
CTGTACCCCAGCTATCATGAACATAGTTCTGTCTCAGCCATGTGGGAAGATCTTTGTAGGGGACCCCCACCAGCAGATCT
ATACCTTCCGAGGTGCAGTGAATGCTCTGTTCACAGTTCCCCACACCCACGTCTTCTATCTCACACAGAGTTTTAGATTT
GGTGTGGAAATAGCCTATGTGGGAGCTACCATTTTGGATGTCTGCAAGAGAGTCAGGAAGAAGACGTTGGTTGGAGGGAA
CCATCAGAGTGGCATAAGAGGTGACATAAAGGGACAAGTAGCTCTGCTGTCCAGGACCAATGCCAATGTGTTTGACGAGG
CTGTCCGGGTAACAGAAGGAGAATCCCCTGCAAGGATACATTTGATTGGGGGGATTAAATCGTTTGGATTGGACAGAATC
ATTGACATTTGGACCCTTCTTCAGCCAGAGGAAGAACGGAGGAAACGAGACCTCATCATTAAAGACAGGTTCATCAGAAG
ATGGGTGCATAAAGAAGGCTTCAGTGGCTTCAAGAGGTATGTGACTGCTGCCGAGGACAAGGAACTGGAAGCGAAGATTG
CGGTAGTTGAAAAATACAATATCAGGATTCCAGAACTGGTAGAAAGGATAGAGAGATGCCACATAGACGATTTGGACTTT
GCAGAGTACATTTTGGGCACCGTGCACAAAGCCAAAGGGTTAGAGTTTGACACTGTGCACGTTTTGGATGATTTCGTGAA
AGTGCCTTGTGCCAGGCATAATCTTGCCCAGCTTCCTCACTTCAGAGTTGAGTCGTTTCTGAGGATGAATGGAATTTAC
TGTATGTTGCTGTTACTCGTGCCAAGAAGCGGCTCATAATGACCAAATCCCTGGAGAACATCCTGACTCTGGCTGGGGAG
TATTTCTTGCAAGCAGAGTTGACAAGTAACGTTTTGAAAACAGGAGTGGTCCACTGCTGTGTGGGGCAGTGCAACAACAC
CATCCCTGTGGACACCATCCTTACCATGAAGAAACTGCCTATCACTTATAGCAACAGGAAGGAAAACAAGGGTGGCTACC
TCTGCCACTCATGTGCAGAGCAGCGCATCGGGCCTTTGGCATTCCTGACTGCCTCCCCAGAGCAGGTGCGTGCCATGGAA
CGCACTGTGGAAGACATAGTGCTGCCCAGGCAGGAGGCCCTGCTCTTCCTTGTCTTCTGAGGCTGTGACTCCACGTGCTA
AAGGAGCCATTCTGGGCTGCATTTCTGCTCACAGGAGACTCTGAGTTCAATCACAGCATCTCTGAGGAGGAAGATAGCAA
CCAGAGTTGGACCTGCCATTTCCCCGTTCCTTACAGAGAGCCAGTTCCCACTTGCCTGTCCTCCAAATGCCTCAGGCCAG
GGAACTGGAAGAAGTTCTTTTGATAAAAAAATCATTTTATGTATTTAAACTTTTATTACAAGGTTTCAATTAAACAGGTA
TTGAGCAGTGGCGTGCCACTGTGAAGTCTAAAAAAAAAAAAAAAAAAAA
```

Figure 8 The complete mouse *UvrD* sequence.

Protein Expression

The computer program DNA Strider indicated that the protein should be 109 kDa in length. The western blot illustrates that the protein may indeed be ~110 kDa in length (see figure 9).

Mouse UvrD Western Blot
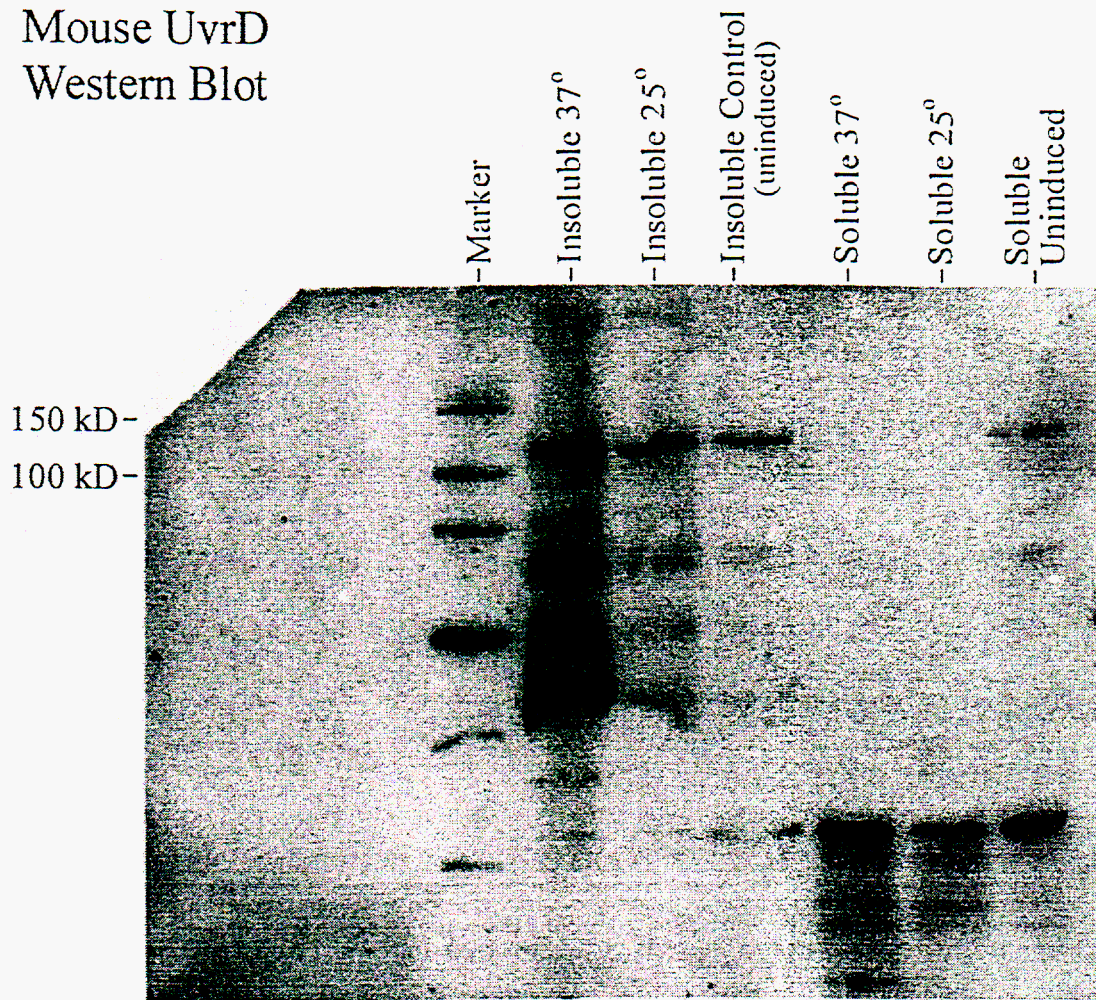
Figure 9 Western blot of protein extracted from bacterial expression cells.

Conclusion

Project Highlights

- Complete mouse *UvrD* sequence has been determined
- Protein has been expressed and is in the process of being purified
- The biochemical function of mouse *UvrD* will soon be investigated

Potential Benefits of Studying Mouse *UvrD*

Human Application

Most of our knowledge about helicases comes from studying bacteria; however, a mammalian model would be more applicable to humans. The mouse protein, UvrD, will be a useful model for the homologous protein found in humans.

Testing Mutations

Further investigation may lead to the discovery that certain mutations in the *UvrD* gene result in faulty DNA repair, replication, or recombination. Studies may also link a defective *UvrD* gene with a disease, such as cancer.

Predicting and Monitoring Disease

Once a defective gene is linked with a disease, researchers will be able to screen humans with mutations in the homologous human *UvrD* gene and thus monitor those having a genetic predisposition for the disease.

References

Snustad, D. Peter, Michael J. Simmons, and John B. Jenkins, *Principles of Genetics,* John
Wiley & Sons, Inc., New York, 1997.

Voet, Donald, Judith G. Voet, Charlotte W. Pratt, *Fundamentals of Biochemistry,* John
Wiley & Sons, Inc., New York, 1999.

Novel Diode-Pumped Solid-State High Average Power Laser Design *

Miguel A. Garcia

University of California, Davis

Lawrence Livermore National Laboratory
Livermore, CA 94551

1/20/00

Prepared in partial fulfillment of the requirements of the Undergraduate Research Semester under the direction of Luis Zapata, Research Mentor, in the Lawrence Livermore National Laboratory.

Novel Diode-Pumped Solid-State High Average Power Laser Design

Miguel A. Garcia, University of California at Davis, Department of Physics

Abstract

A proprietary laser design uses diode pump-light introduced at the edges. Using ray trace models, we investigated the optical geometry and optimized dimensions to balance the efficiency and pump uniformity. Our design yielded an optimized source distribution for a 6 mm aperture disk. Experimental results were in good agreement with the predictions of the computer model.

Introduction

State of the art solid state laser systems of today use diode lasers as the pumping source to produce efficient, high-power lasers of tens of watts for industrial and scientific applications. This generation of novel lasers is compact, and versatile, finding applications ranging from the laboratory to the machine shop. Conventional flashlamps are replaced by diode arrays allowing for greater pump power densities[1]. In addition, clever pump guiding methods are used to optimize the irradiance of the laser medium [1,2]. The use of Yb:YAG thin disks, diode-pumping, and skillful pumping techniques make it feasible to create a compact high power efficient laser.

1 System structure

The design of this novel diode pumped solid-state high power laser system (Figure 1) has the goal of maximizing the efficiency, quality, and power of the output beam. Our methodology invokes the excellent lasing properties of Yb:YAG and the application of novel heat removing techniques. Furthermore, the main idea behind our approach is to mount one face of a very thin Yb:YAG crystal disc on a heat sink.
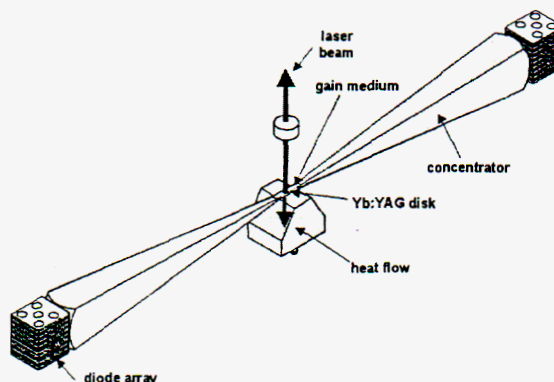
*Figure 1. Three-dimensional view of the Yb:YAG thin disk crystal being side pumped with 940 nm diode laser arrays through concentrators. This figure has been reviewed and released, **ucrl-vg-136440**.*

2 The thin disk concept

Production of high power laser light from compact devices is problematic. Excess heat in the laser medium causes stress birefringence (de-polarization) and thermal lensing [1,3,4]. The thin disk geometry (Figure 2) allows for effective heat removal where the heat flux is one dimensional and longitudinal to the output beam, and thus the quality of the output beam is dramatically improved.
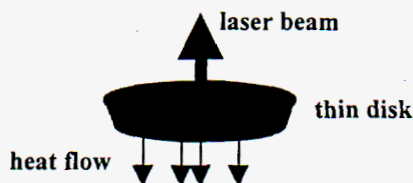
Figure 2. Thin disk principle. Longitudinal heat flow does not distort output beam, and the power handling capability scales as 1/thickness.

3 TracePro modeling

A ray-trace model of this proprietary laser design has been developed to investigate the effect of various

dimensions on the efficiency and source distribution. This model has been excersized in order to optimize the specific geometry of the active medium, diode pump configuration, and operating conditions.

3.1 Diode array modeling

Far Field measurements characterizing diode array divergence were used in the TracePro ray tracing program. The quasi-three level nature of the Yb ion requires a high pump intensity (~10-50 Kw/cm^2). The limit of concentration depends on the divergence properties of the diode array. Source distribution and diode light transport efficiency were realistically computed. For further operational descriptions of the diode array see [4].
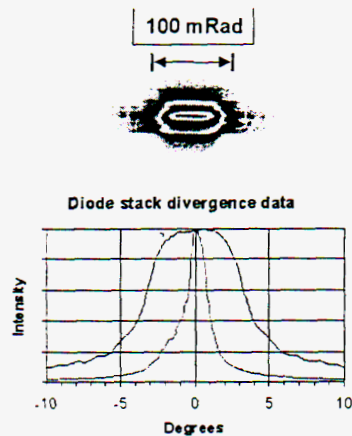
Figure 3. Image showing the measured divergence of a 40-bar diode array. The characteristic divergence angles are approximately: horizontal = 8° and vertical = 1.5°.

3.2 Diode Pumping

The requirement of operating the proprietary laser design dictates the use of specially designed lens-ducts of high transfer efficiency. The transfer efficiency was measured for the lens-ducts used. Figure 4 is graph of the efficiency transfer for one lens-duct.

A lens-duct concentrates the light of a diode laser array and conditions it to enter through the edge of the gain medium. About 20 Kw/cm^2 of 940 nm diode pump light is coupled into the numerical aperture of the proprietary design. The predicted electrical-to-optical efficiency for this laser system is ~23% at 580 watts output of 1030 nm light.
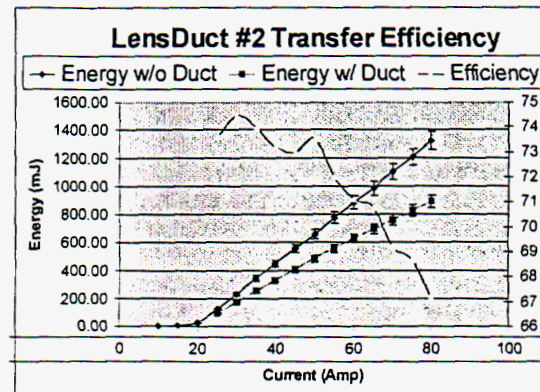
Figure 4. Transfer efficiency chart of a of the two lens-ducts used to pump the Yb:YAG disk.

Figure 5 shows a schematic view of the path the diode light (principal rays) is subjected to for the purpose of achieving the high pumping density needed to raise the Yb^{3+} ions above laser threshold. The lens-duct concentrates the light and it passes the active medium multiple times. This means that the local threshold pump power density is absorbed efficiently for cw operation [5]. Figure 5 is an augmented view of the disk section of the computer model shown in Figure 6.
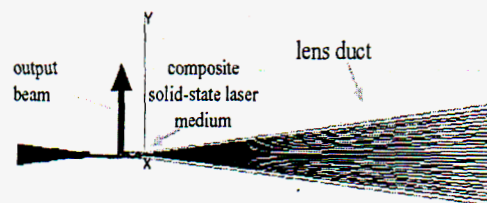
Figure 5. Diagram of principal rays concentrated by the lens-duct to enter the Yb:YAG medium for absorption.

In the computer model (Figure 6), accepted values for the optical properties of the composite medium were used to

simulate our laser system. Light sources with similar characteristics to the real diode light and analyzed for optimal absorption by the laser medium. The absorption coefficient and refraction index used were 1.5 per mm and 1.8315 respectively.

Figure 6. Schematic view of TracePro computer model.

Results

The source distribution was mapped (Figure 7) for 6 mm and 12 mm aperture crystals. TracePro modeling predicts a trade-off between pumping efficiency and pumping homogeneity. The short disk presents a uniform distribution, but has only an 82% optical-to-optical efficiency. In contrast, the long disc presents a non-uniform distribution, and it yields a 90% optical-to-optical efficiency.
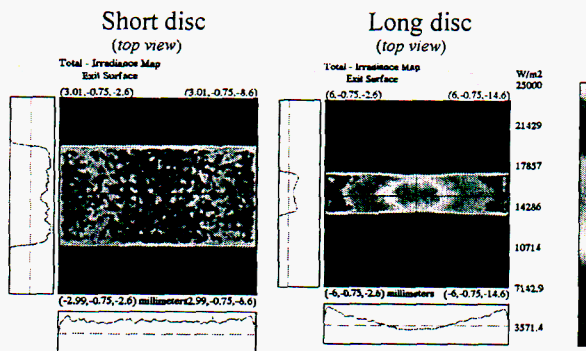
Figure 7. Source distribution of 6 mm and 12 mm thin disc crystals.

Conclusion

On going experiments with first generation prototypes are in good agreement with theoretical predictions of computer model design. Future work includes fluorescence measurements and output beam imaging.

References

1. Giesen, A.; Seeber, W.; Keller, U. *Ultrafast ytterbium-doped bulk lasers and laser amplifiers.* Applied Physics B (Lasers and Optics) Appl. Phys. B, Lasers Opt. (Germany), vol.B69, (no.1), Springer-Verlag, July 1999. p.3-17. 79 .

2. Giesen, A.; Hugel, H.; Voss, A.; Wittig, K.; Brauch, T.; Opower, H. *Scalable concept for diode-pumped high-power solid-state lasers.* Applied Physics B (Lasers and Optics) Appl. Phys. B, Lasers Opt. (Germany), vol.B58, (no.5), May 1994. p.365-72. 4 references.

3. Brauch, U.; Contag, K.; Giesen, A.; Johannsen, I.; Karszewski, M.; Schiegg, U.; Stewen, C.; Voss, A. *Multi-hundred-watt diode pumped Yb:YAG thin disc laser. Proceedings of the SPIE-* The International Society for Optical Engineering Proc. SPIE - Int. Soc. Opt. Eng. (USA), vol.2986, (Solid State Lasers VI, San Jose, CA, USA, 10-11 Feb. 1997.) SPIE-Int. Soc. Opt. Eng, 1997. p.2-9.

4. Beach, R. J.; Dane, B; Hackel, L. A.; Zapata, L; *Laser driver for soft-x-ray projection lithography.* Applied Optics, Vol. 32, No. 34. Dec. 1993.p.1614-6919.

5. Contag, K; Karszewski, M; Stewen, C Giesen, A; Hugel, H; *Theoretical modelling and experimental investigations of the diode-pumped thin-disk Yb : YAG laser.* Quantum Electronics; AUG 29 1999.p.697-703.

Crystallization of Bacterial Toxins for

X-ray Crystallographic Analysis

By Sadia S. Khan

Bernhard Rupp, Macromolecular Crystallography Group
Biology and Biotechnology Research Program,
Lawrence Livermore National Laboratory
Office of Defense Programs, Department of Energy

Submitted in partial requirements for the Undergraduate Research
Semester Program

Monday, December 13, 1999

Abstract

The purpose of this study is to find suitable crystallization conditions for two bacterial toxins, mutant *Staphylococcus aureus* enterotoxin B (mutant SEB) and *Clostridium botulinum* neurotoxin A light chain (BoNT A LC), and for a selenomethionine derivative of BoNT A LC. Our anticipated goal is to produce crystals amenable to X-ray crystallographic analysis. Initial crystallization conditions were found through random screens and refined in further experiments to improve crystal quality. Crystals were grown by hanging drop vapor diffusion at 4°C.

Initial results indicate that:

- For mutant SEB:
    - Ethanol crystallized the protein at low and high ethanol concentrations as well as at low and high buffer pH. The results of the initial hits, however, could not be repeated nor refined successfully. An unexpected variable may have been the type of sealant used to isolate the well-drop system from the surroundings. Oil was used as the sealant in most of the random screens whereas grease or lubricant was the sealant in the repeat and refinement experiments. In future experiments, all successful random conditions will be repeated, with the sealant as the variable, to determine how the type of sealant affects crystallization for each random condition.

    - Mg^{2+} or Cd^{2+} in combination with PEG 10K yielded small crystals and spherulites, respectively. Both metal and PEG 10K concentrations will be varied in future experiments to improve crystal quality and size. In addition, Mg^{2+} in combination with HEPES buffer (pH 8.5) yielded crystals. The effects of various buffers and buffer concentrations in combination with Mg^{2+} will be tested.

- For BoNT A LC:
    - A specific combination of PEG MME 5K and Na succinate buffer (pH 4.5) yielded needle-shaped crystals. The results of this initial hit, however, could not be repeated nor refined successfully. An unexpected variable may have been the age of the protein used in the crystallization experiments. The protein used in the random screen was purified about 15 days prior to the experiment, whereas the protein used in the repeat experiments was purified about 39 days prior.

    - A specific combination of $(NH_4)_2SO_4$, Na acetate buffer (pH 6.5), and glycerol yielded spherulites. The concentrations of the reagents and buffer pH will be varied in further experiments.

    - Crystallization of the Se-Met derivative was not successful. Additional random conditions will be screened.

Introduction

Over the past decade, interest in toxins has grown due to their potential threat as biological weapons, both on the battlefield and in isolated attacks of terrorism. Toxins are most often proteins, synthesized and secreted by several species of fungi, bacteria, and viruses. Toxins disrupt human cellular processes, leading to illness or even death. Like other proteins, toxins are defined by three levels of structure. The most fundamental level is the primary structure, which is simply the sequence of amino acids in the protein. The secondary structure is the local conformation of the polypeptide backbone, which arises from hydrogen bonds formed between nonadjacent peptide bonds. Interest in toxins, however, is centered on their tertiary structure. The tertiary structure is the overall shape or conformation of the protein, which arises from arrangement of secondary structural motifs and complex interactions among the side chains of the amino acid residues. For monomeric proteins, tertiary structure dictates protein function; therefore, alterations of tertiary structure past some level of tolerance disrupt protein function. By determining the tertiary structure of toxins, we can understand the physical mechanism of diseases caused by the toxins, and we can design effective inhibitors and/or vaccines against them. The tertiary structure of each protein is the direct result of its unique primary structure, but no theory exists which allows reliable prediction of tertiary structure from primary structure. Therefore, empirical methods, such as X-ray crystallography and NMR, are needed to solve toxin structure.

Two toxins of interest are bacterial in origin. *Staphylococcus aureus* enterotoxin B (SEB) causes food poisoning. If left untreated, the illness can be fatal. SEB belongs to a whole class of related enterotoxins (types A, C1, C2, C3, D and E). SEB is classified as a superantigen because it causes the proliferation of a wide subset of T cells. SEB functions at the molecular level by forming a complex with two proteins of the immune system, a major histocompatibility complex II (MHC II) protein and a T cell receptor. Complex formation triggers proliferation of T cells, which leads to increased production of cytokines, the inflammatory factors responsible for the stomach cramps and fever associated with the illness. In recent studies, a mutant form of SEB was found not to cause illness, but it did raise antibodies against the native toxin (Bavari, 1996 and Ulrich 1995). The mutant binds the T cell receptor, but it cannot bind MHC II. The structure of native SEB, alone and in complex with a MHC II protein, has already be   ed (Papageorgiou, 1998 and Jardetzky, 1994). To understand how the mutatio..s a..   ..c structure of SEB and as a final step in evaluating the safe use of this mutant as a vaccine, we will solve the mutant SEB structure by X-ray crystallography.

Another toxin of interest is *Clostridium* botulinum neurotoxin A (BoNT A), which is the causative agent of botulism. This toxin belongs to a whole family of related clostridial neurotoxins, which includes tetanus toxin and the other botulinum toxins (types B, C1, C2, D, E, F, and G). BoNT A is one of the most common botulinum neurotoxins and is the most harmful to humans. At the molecular level, the toxin functions at the neuromuscular junction (NMJ). BoNT A binds to receptors on the nerve cell surface, and is then internalized by receptor-mediated endocytosis. Once inside the cytoplasm, BoNT A cleaves SNAP-25, one of the key proteins involved in neurotransmitter release. This

leads to blockage of neurotransmitter release at the NMJ, which causes paralysis of the adjacent muscle. Native BoNT A is composed of two polypeptide chains held together by a sulfur bridge. One chain is 100 kDa and is known as the heavy chain, whereas the other chain is only 50 kDa and is dubbed the light chain. The heavy chain is responsible for internalizing the toxin into the nerve cell. The light chain is the catalytic fragment of BoNT A; it is a Zn^{2+}-dependent endopeptidase that cleaves the target protein. However, once inside the cell, the light chain must separate from the rest of the toxin before becoming active. The structure of the native toxin is already known (Lacy, 1998). To understand how the structure of the light chain changes after separation, we will solve the light chain structure alone by X-ray crystallography.

In X-ray crystallography, proteins are analyzed as single crystals. Protein crystallization, however, is currently the bottleneck in crystallography. Unlike salt ions, proteins are much more difficult to crystallize for four reasons. First, protein molecules are held together in the lattice by weaker, noncovalent hydrogen bonds whereas ions are held together by strong ionic bonds. Second, proteins are essentially large, flexible polymers composed of thousands of atoms; therefore, it is difficult to pack proteins in a highly ordered state. Third, protein crystals are, on average, composed of 50-70% solvent by volume. Although the benefit of this is that the protein remains hydrated and is most likely in its native form, the resulting crystal is prone to disorder. Fourth, no theory exists that can predict under what conditions each type of protein can crystallize. Each protein behaves differently with regard to crystallization; thus, suitable crystallization conditions must be found for each protein empirically.

The goals of this study were to find suitable crystallization conditions for mutant *Staphylococcus aureus* enterotoxin B (mutant SEB) and *Clostridium* botulinum neurotoxin A light chain (BoNT A LC). In addition, crystallization conditions were also screened for a selenomethionine (Se-Met) derivative of BoNT A LC for crystallographic multiwavelength anomalous dispersion (MAD) experiments. Suitable crystallization conditions are defined as conditions that will yield single crystals that are large (0.5mm in the smallest dimension), and three-dimensional with well-defined edges. For mutant SEB and BoNT A LC, initial conditions were found through random screens, and then refined to improve crystal quality. However, further work is still needed to find conditions that will yield crystals suitable for X-ray crystallographic analysis.

Materials and Methods

Mutant SEB was kindly provided by R.G. Ulrich (Army Research Institute in Maryland). BoNT A LC and its selenomethionine (Se-Met) derivative were synthesized as recombinant proteins and were kindly provided by Rod Balhorn (Lawrence Livermore National Laboratory).

Crystallization Strategy

First, we screened for conditions that would yield any kind of crystals. This was done by testing the Hampton crystal screens, and by random sampling of different reservoir solutions. The Hampton crystal screens were available commercially (Hampton Research). To make up the random sample, we used Crystool, a software program developed by Segelke et al., to generate 9 random screens, with each screen containing 48 different reservoir solutions. The reservoir solutions were random combinations of commonly used precipitants and co-precipitants, biological buffers, pH, and detergents (see Table 1). 24 experiments were carried out at a time in 24-well microtitre plates (see Figure 1). Mutant SEB was screened with two Hampton crystal screens and with three random screens, whereas BoNT A LC was screened with four random screens (see Table 2). The Se-Met derivative of BoNT A was screened with two random screens. No crystals were produced from the Hampton crystal screens. Some of the random screen conditions yielded crystals for mutant SEB and BoNT A LC, but these were too small and/or of poor quality. We then varied one or more of the parameters of the initial success in hopes of improving crystal quality and size.

Figure 1. All crystallization experiments were carried out in 24-well microtitre plates such as the one shown below.
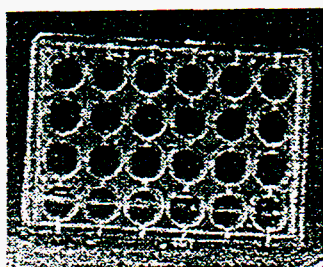
Crystallization Method

We crystallized the protein using the hanging drop vapor diffusion method at 4°C because this procedure gradually crystallizes the protein without denaturing it (see Figure 2). There is a small chamber, or well, which contains 600 or 1000 µl of the reservoir. The drop that is mounted over the reservoir is composed of protein and reservoir, both present in undersaturated states. The coverslip seals off the well and the drop from the surroundings. Under these conditions, diffusion of water vapor and volatile components occurs between the drop and well only. Because the concentration of the reservoir in the drop is lower than in the well and because the well volume is much larger than the drop volume, net diffusion occurs so that the reservoir concentration in the drop increases to approach the well concentration and the drop volume decreases. This, in turn, slowly

increases the protein concentration in the drop to supersaturation, where crystallization can occur.

Figure 2. Crystals were grown by hanging drop vapor diffusion at 4°C.
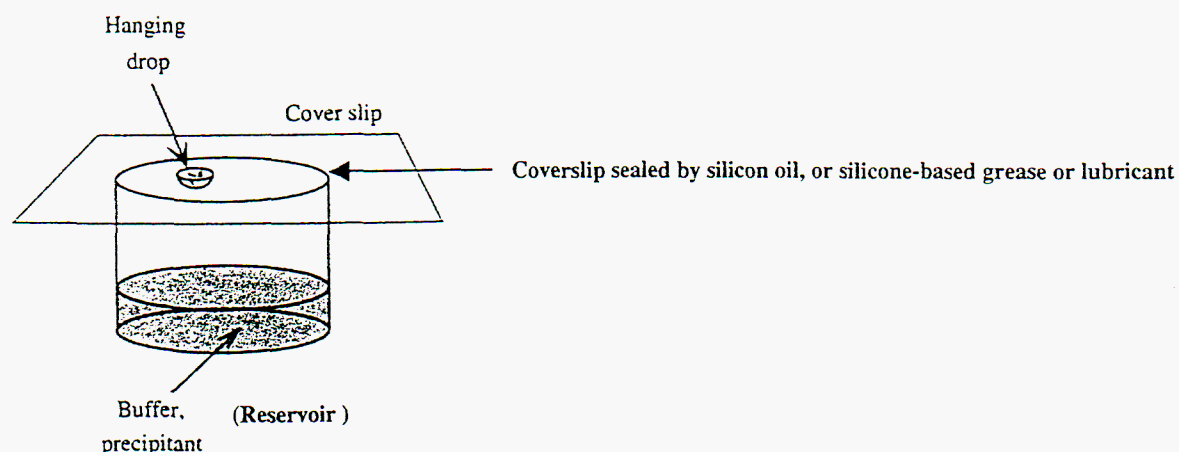
For most Hampton crystal screens, random screens, and refinement and repeat experiments without microseeding:
Hanging drop = 50% protein + 50% reservoir

For refinement and repeat experiments with microseeding:
Hanging drop = 13 or 17% protein + 87 or 83% reservoir

SDS-polyacrylamide gel electrophoresis

SDS-PAGE was used to detect degradation products of BoNT A LC. 45-µl aliquots of BoNT A LC, 6.0 mg/ml, were stored at room temperature, 10°C, 15°C, and 4°C 21 days after purification. Samples of BoNT A LC from each temperature were then analyzed by SDS-PAGE 35 days after purification. For each temperature, 3 µl of BoNT A LC was dissolved in 15 µl 1x MES running buffer (Novex), followed by heating for 10 minutes in a hot H_2O bath and centrifugation for 4 minutes at 5000 rpm at 4°C. The samples were run on a 4-12% bis-Tris gel (Novex) for 36 minutes at 200 V, and visualized by Coomassie blue stain. 20 µl of MW markers (Novex Multi-Mark Multi-Colored) were also loaded and run on the gel for comparison.

Results and Discussion

Mutant SEB Crystallization

Ethanol (EtOH) and the divalent cations, Mg^{2+} and Cd^{2+}, were observed to crystallize mutant SEB.

Initial Hits

In the random screens, three reservoir solutions containing EtOH as the major precipitant yielded crystals (see Table 3). Overall, mutant SEB crystallized over a wide range of EtOH concentrations (15-57% by volume), and over a broad range of pH (5.5-8.5). The initial crystals, however, were not suitable for X-ray diffraction analysis (see Figure 3). The crystals produced from R8 #32 were small (0.2-0.3mm est.), roughly formed, and submerged in precipitate. The crystals produced from R6 #43 were also small (0.2-0.3mm est.), and twinned. The single crystal produced from R9 #25 was the most promising because it was larger (0.4mm est.) with well-defined edges, but it had not sufficiently grown in one dimension.

In the random screens, two reservoir solutions containing Mg^{2+} as the major and co-precipitant, respectively, yielded crystals (see Table 3). R5 #1 yielded two microcrystal clusters, whereas R5 #10 yielded needle-shaped crystals emanating from a single nucleation site (data not shown). Neither crystal form was acceptable for analysis. The microcrystal clusters were too small, and the needle-shaped crystals were one-dimensional.

Refinement

One or more of the parameters of the reservoir solutions were varied to improve crystal quality and size.

Ethanol Series

For the EtOH containing solutions, the parameters of R8 #32 and R9 #25 were varied with little success. For R8 #32, 48 experiments varying EtOH concentration, buffer, pH, type of secondary alcohol, and mutant SEB stock concentration, were conducted (see Figure 3 and Table 4a). The concentration of buffer and secondary alcohol remained the same. All of the drops precipitated, but no crystals were observed. For R9 #25, 24 experiments varying EtOH concentration and type of buffer were conducted. The concentration and pH of the buffers were not varied from the original random solution, except in the first six experiments, in which CAPSO concentration was 0.04M in the reservoir. Again, all the drops precipitated but no crystals were observed. Because the crystal obtained from R9 #25 was the most promising out of the initial hits, R9 #25 conditions were repeated in 6 experiments. All the drops precipitated but no crystals were observed. In contrast, the initial crystals from R9 #25 appeared 6 days into the experiment. This indicates some other variable was inadvertently affecting the Random 9 experiment. As noted in Table 3, oil was used to seal the coverslips in Random screens 8

and 9, whereas grease or lubricant was used to seal the slips in all other experiments. Oil was replaced with grease or lubricant as the sealant in later experiments because oil did not provide an airtight seal between the well-drop system and the surroundings. This was evident by the accumulation of debris in many of the drops in Random screens 8 and 9. Therefore, the conditions in the random screens may not have been identical to the conditions in later experiments. Some of the components may have partly evaporated from the well and/or drop in the random screens, thus changing the concentration of the components in the initial conditions. Or, a piece of debris could have acted as a nucleation site, thus allowing crystals to form in the initial conditions.

Divalent Cation Series

We refined the parameters of the R5 #1 reservoir solution with greater success (see Table 4b). The appearance of microcrystals in the initial hit indicated multiple nucleation sites. Therefore, the crystals of R5 #1 were used to microseed other experiments to obtain a few, large single crystals. The conditions of R5#1 were repeated in 6 microseeding experiments. However, these experiments produced only microcrystals as well (see Figure 3). To further improve crystal quality and size, these microcrystals were in turn used to seed 9 experiments varying Mg^{2+} over low concentration (2-10 mM). However, these conditions did not produce any crystals at all.

The literature indicates that divalent metal cations help proteins crystallize by promoting contacts between protein molecules in the crystal lattice (Trakhanov, 1995 and 1998). Therefore, various divalent cations were tested. The type and concentration of divalent cations were varied in a series of 18 experiments conducted without microseeding. The metals tested were Ca^{2+} (2-7mM), and Co^{2+} and Ni^{2+} (20-70mM). No crystals were produced. The type and concentration of divalent cations were also varied in a series of 12 experiments conducted with microseeding. The metals tested were Ca^{2+} and Cd^{2+} (200-700mM). Microcrytals from the R5 #1 initial hit were used as the seed. $[Cd^{2+}]$ = 300-700mM yielded brown spherulites. Cd^{2+} concentrations were further varied (500-1400 mM) in a series of 15 experiments, but these concentrations also yielded spherulites. Furthermore, the spherulites are not stable and appear to undergo a bizarre transformation (see Figure 4). Four days into the experiment, the spheres were clustered together. Ten days into the experiment, the characteristic brown color of the spheres was fading, and brown clumps of matter were observed as separate entities in the drop. By 52 days, the spheres were almost completely transparent and the brown clumps got larger. Therefore, the brown material may have leaked out from the spheres and coalesced.

BoNT A LC Crystallization

Initial hits

Only specific combinations of reagents yielded crystals for BoNT A LC (see Table 5). R6 #16 yielded spherulites submerged in brown precipitate, and R6 #18 yielded needle-shaped crystals (see Figure 5). Neither result was suitable for X-ray diffraction analysis. The spherulites were small (0.2-0.3mm est.) and lacked well-defined edges, whereas the needles were one-dimensional. No initial hits were observed for the Se-Met derivative of BoNT A LC.

Refinement

The parameters of R6 #18 were varied with little success. 18 experiments were conducted varying PEG MME 5K concentrations and buffer pH (see Table 6). Successful crystallization of proteins with a ligand or inhibitor have been reported (Bergfors, 1999), so BoNT A LC was also crystallized in complex with two inhibitors provided by Jim Schmidt (USAMRIID), but with little success.

Six experiments repeating the conditions of R6 #18 were also conducted, but no crystals appeared. The protein used in random screen 6 was purified about 15 days prior to the experiment, whereas the protein used in the repeat experiments was purified about 39 days prior. The protein may degrade over time, which could have caused differences in crystallization behavior. To look for evidence of protein degradation, BoNT A LC was analyzed by SDS-PAGE 35 days after purification. 45-µl aliquots of BoNT A LC, 6.0 mg/ml, were stored at room temperature, 10°C, 15°C, and 4°C 21 days after purification. Samples of BoNT A LC from each temperature were then analyzed by SDS-PAGE 35 days after purification (see Figure 6). A prominent band was present at 52 kDa for all temperatures, which indicates that the light chain was the predominant species. A second band at around 98 kDa was also present for all temperatures, which indicates that a high-molecular weight contaminant is present in the samples. Furthermore, several bands below 52 kDa were present for all temperatures. The bands were more intense as temperature increased. This indicates that BoNT A LC may be undergoing degradation, and that the rate increases with rise in temperature. Therefore, the advanced age of the BoNTA LC sample used in the repeat experiments could explain why no crystals appeared. The protein used in the PEG MME 5K vs. buffer pH experiment (see preceding paragraph and Table 6) was also purified 39 days prior to the experiment, and therefore, may explain why no crystals appeared.

Conclusions

Mutant SEB Crystallization

Ethanol Series

In the random screens, EtOH crystallized mutant SEB over a broad range of EtOH concentration and buffer pH. An unexpected variable that may have affected the crystallization, however, is the type of sealant used (oil vs. lubricant vs. grease) because the initial success of R9 #25 could not be repeated when grease, and not oil, was used as the sealant. Furthermore, none of the refinement experiments for R8 #32 and R9 #25 yielded crystals, although a wide range of EtOH concentration and buffer pH were tested. In the refinement experiments, either grease or lubricant was used as the sealant, whereas oil was the sealant in Random screens 8 and 9. In future experiments, all successful random conditions will be repeated, with the sealant as the variable, to determine how the type of sealant affects crystallization for each random condition.

Divalent Cation Series

In the random screens, Mg^{2+} crystallized mutant SEB at two different Mg^{2+} concentrations, 1.0M and 0.14M. In refinement experiments of R5 #1, low Mg^{2+} concentrations (2-10mM) did not yield crystals. Experiments refining R5 #1 conditions at higher Mg^{2+} concentrations (50-140mM) are in progress. Out of various other divalent cations tested, only Cd^{2+} yielded positive, albeit uncommon, results. In future experiments, lower Cd^{2+} concentrations will be tested. Also, still other divalent cations will be tested, such as Mn^{2+}, Zn^{2+}, etc. PEG 10K concentrations will also be varied. Lastly, the conditions of R5 #10 (see Table 3) will be refined. In particular, the effects of various buffers and buffer concentrations in combination with Mg^{2+} will be tested.

BoNT A LC Crystallization

Only specific combinations of reagents yielded crystals of BoNT A LC (see Table 5). The conditions of R6 #18 were repeated and refined without success, however. This could be due to the advanced age of the protein used in the crystallization experiments. In future experiments, the degradation of BoNT A LC will be monitored over time. In the meantime, the repeat and refinement experiments of R6 #18 will be carried out using fresh protein. The conditions of R6 #16 will be refined also using fresh protein (see Table 5). In particular, the concentrations of the reagents and buffer pH will be varied in further experiments. Furthermore, additional random conditions for the Se-Met derivative will be screened.

Figure 3. Sample of results from mutant SEB crystallization experiments

(experiments = black, results = red)

Ethanol Series

Random 8 #32   Random 6 #43   Random 9 #25
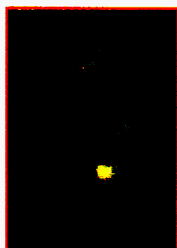
105.6x   105.6x   105.6x

%EtOH vs. change in pH,
isopropanol/sec-butanol
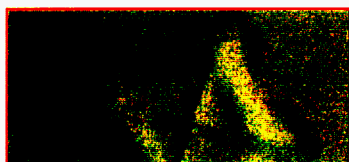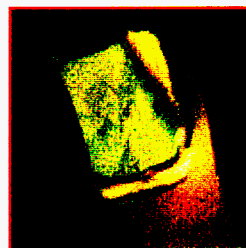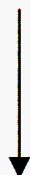
%EtOH vs. change in buffer
+
Repeat R9 #25

Figure 3. Sample of results from mutant SEB crystallization experiments

Divalent Cation Series

Random 5 #1

Repeat R5 #1 (+) microseed

80x

$[Mg^{2+}]$ = 2-10 mM (+ microseed)

Divalent metals
(+) microseed
Ca^{2+}, Cd^{2+}
(-)microseed
Ca^{2+}, Co^{2+}, Ni^{2+}

$[Cd^{2+}]$ = 300-500 mM

$[Cd^{2+}]$ = 500 – 1400 mM

57.6x
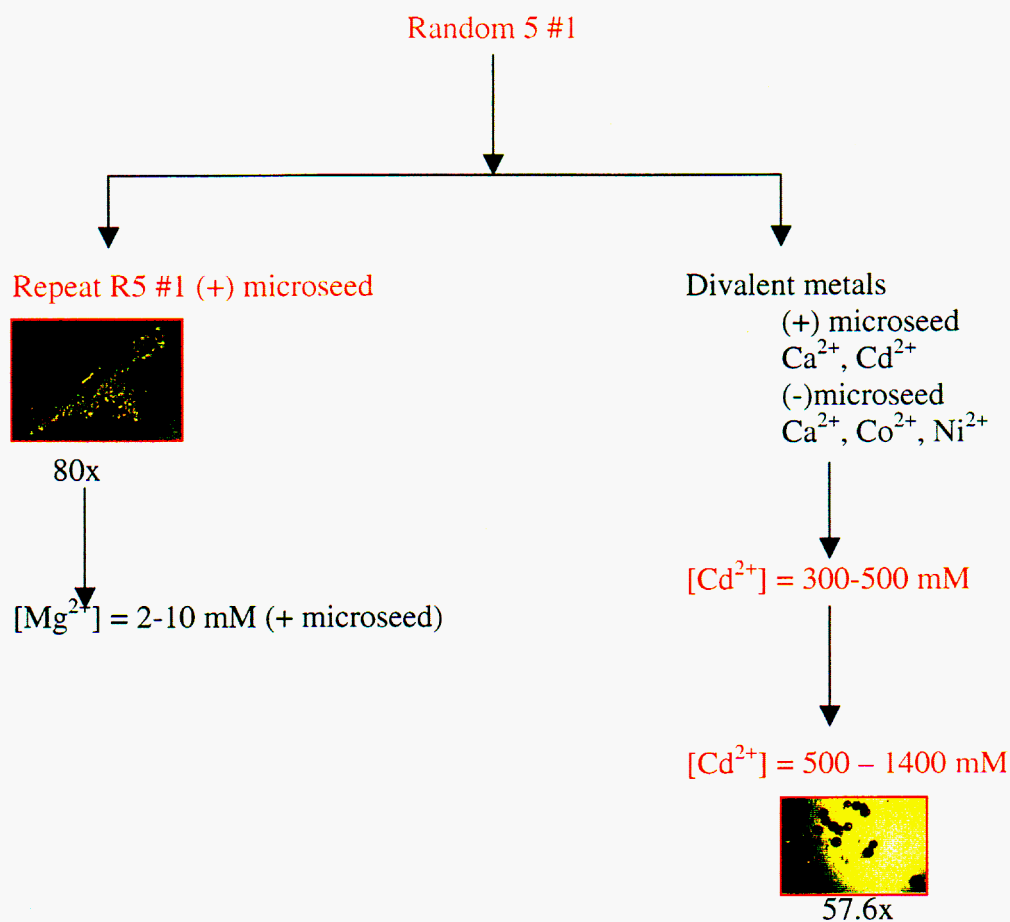
Figure 4. Picture of spherulites 11 days into the $[CdCl_2]$ experiment
(see Table 4b). The spheres were losing their characteristic brown color, and brown
clumps of matter were observed as separate entities in the drop.
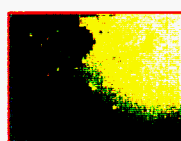
105.6x

Figure 5. Samples of results from BoNT A LC crystallization experiments

(experiments = black, results = red)

Random 6 #16
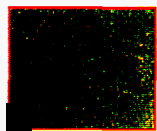
80x

Random 6 #18
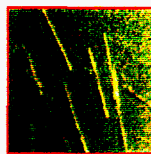
105.6x

%PEG MME 5K vs. change in buffer pH

+

Repeat R6 #18

High MW contaminant

Prominent band = light chain

Low MW degradation
fragments of light chain

185 kDa
98

52
31

19/17
11
6
3

Low MW degradation
fragments of light chain

4°C   10°C   MW   15°C RT*

*RT = Room Temperature
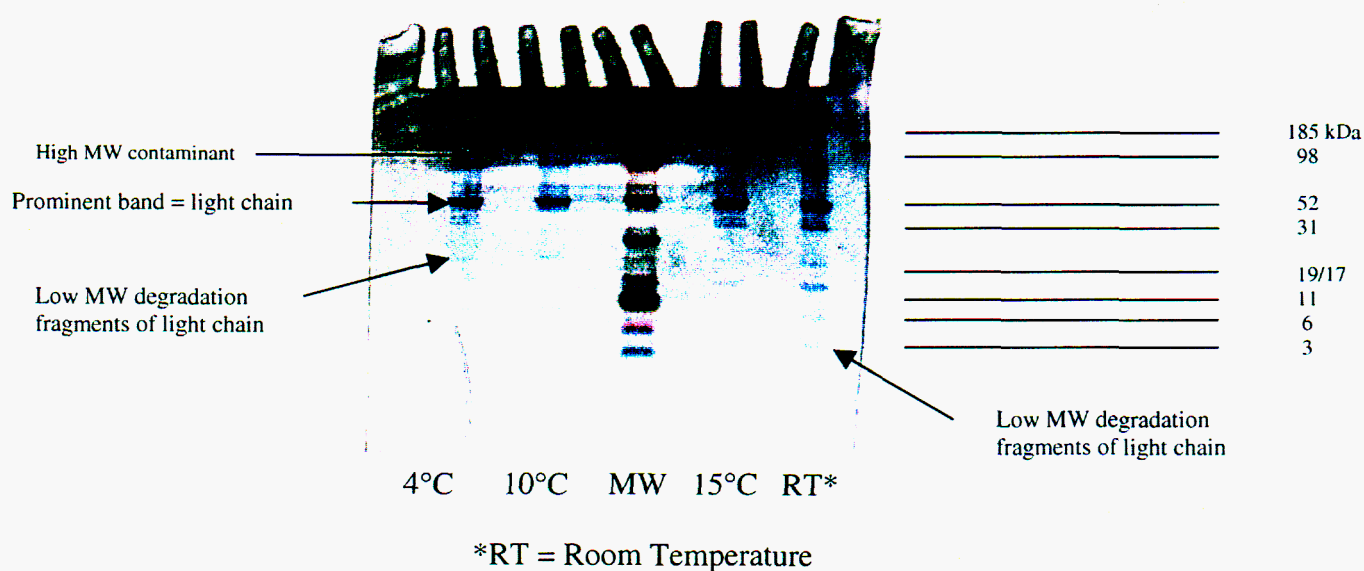
Figure 6. Degradation analysis of BoNT A LC. Coomassie blue-stained SDS-polyacrylamide gel of BoNT A LC from samples stored at 4 different temperatures. 1 µg of each temperature sample was loaded. Numbers to the far right indicate MW of corresponding bands in the MW lane.

Table 1. Common precipitants, co-precipitants, biological buffers, and detergents used to make up the random screens.

Precipitant and/or Co-Precipitant

Salts

1	NaK tartrate
2	NaK phosphate
3	Na acetate
4	Na formate
5	Na citrate
6	NaCl
7	KCl
8	Mg formate
9	MgCl$_2$
10	MgSO$_4$
11	Mg acetate
12	LiSO$_4$
13	Ammonium phosphate
14	(NH$_4$)$_2$SO$_4$
15	Ammonium acetate
16	Ammonium formate
17	CaCl2
18	Ca acetate
19	MnCl2
20	Zn acetate

Polyethylene glycols (PEG)

1	PEG 400
2	PEG 2K
3	PEG 4K
4	PEG 6K
5	PEG 10K
6	PEG-MME* 750
7	PEG-MME 2K
8	PEG-MME 5K
9	PEG-DME* 2K

Alcohols

1	ethanol
2	2-butanol
3	isopropanol
4	MPD
5	Methanol
6	Hexanediol

Others

1	DMSO
2	EDTA

*MME = monomethyl ether
*DME = dimethyl ether

Biological buffers

		pH range	
		minimum	maximum
1	Na acetate	4.5	6.5
2	HEPES	6.5	8.5
3	Na cacodylate	4.5	7.5
4	Na citrate	4.5	7.5
5	Na succinate	4.5	7.5
6	NaK phosphate	5.5	8.5
7	Tris HCl	6.5	8.5
8	Tris maleate	4.5	6.5
9	Imidazole maleate	4.5	8.5
10	BisTrisPro	5.5	7.5
11	CAPSO	8.5	9.5

Detergents

1	glycerol
2	n-nonyl beta-D-glucopyranoside
3	n-octyl beta-D-glucopyranoside
4	octyl beta-D-thioglucopyranoside
5	heptanoyl-N-methyl glucamide
6	nonanoyl-N-methyl glucamide
7	octanoyl-N-methyl glucamide
8	lauryl dimethylamine oxide

Table 2. Initial screening.

Mutant SEB was screened using both Hampton crystal screens and random screens,
whereas BoNT A LC and the Se-Met derivative were screened through random screens only.
Hampton crystal screens were available commercially, but the random screens were generated using Crystool, a software developed by Segelke et al.

Mutant SEB	BoNT A LC	Se-Met derivative of BoNT A LC
Hampton Crystal Screens I, II		
Random Screens 5, 6 (#35and #43), 8, 9	Random Screens 1, 2, 3, 6	Random Screens 1, 2 (partial)

Table 3. For SEB: the composition of the successful reservoir solutions from the random screens.

Ethanol Series

	R6 #43	R8 #32	R9 #25
Random screen, reservoir solution #	R6 #43	R8 #32	R9 #25
Precipitant	14.9% Etoh	23.3% Etoh	5.% Etoh
Buffer	0.1 M Na cacodylate	0.1 M NaK phosphate	0.1 M CAPSO
pH	7.5	5.5	8.5
Co-precipitant	0.04 M $(NH_4)_2SO_4$	9.5% Isopropanol	
Detergent			4.4% glycerol
Sealant	grease	oil	oil
# of days before crystals appeared	7	9	6

Divalent cation series

	R5 #1	R5 #10
Random screen, reservoir solution #	R5 #1	R5 #10
Precipitant	5.5% PEG 10K	1.0 M $MgSO_4$
Buffer		0.1 M HEPES
pH		8.5
Co-precipitant	0.14 M $MgCl_2$	0.37 M NaCl
Detergent		4.3% glycerol
Sealant	grease	grease
# of days before crystals appeared	7	7

43

Table 4a. Summary of refinement experiments for SEB: ethanol series.

1 Change in Volume% Ethanol vs. change in type of buffer

change in buffer (well concentration)   %EtOH ⎯⎯⎯⟶

	20	30	40	50	60	70
0.04 M CAPSO	1	2	3	4	5	6
0.1M HEPES	7	8	9	10	11	12
0.1M Tris-HCl	13	14	15	16	17	18
0.1M imidazole maleate	19	20	21	22	23	24

Stock [SEB] (mg/ml)   12.12
Drop composition (protein:well), microliters   2:02
Well volume (microliters)   1000
Sealant   grease
Constant in all wells   4.4% by volume glycerol, pH 8.5
Origin   R9 #25

2 Change in Volume% Ethanol vs. change in pH and secondary alcohol

%EtOH and change in secondary alcohol ⎯⎯⎯⟶

	20%, isopropanol	20%, secbutanol	40%, isopropanol	40%, secbutanol	60%, secbutanol	60%, isopropanol
5.5	1	2	3	4	5	6
6.5	7	8	9	10	11	12
7.5	13	14	15	16	17	18
8.5	19	20	21	22	23	24

change in pH

Stock [SEB] (mg/ml)   12.12
Drop composition (protein:well), microliters   2:04   5
                                                                              2:02
Well volume (microliters)   1000   1000
Sealant   lubricant   lubricant
Constant in all wells   10% by volume secondary alcohol, 0.1M NaK phosphate buffer
Origin   R8 #32

Table 4b. Summary of refinement experiments for SEB: divalent cation series. (Note: X indicates positive result)

1   Change in [$MgCl_2$] (+ microseed)

change in divalent cation

[Divalent cation] (mM) →

2mM	3	4	5	6	7
1	2	3	4		6

8 mM

7	8	9	10

Stock [SEB] (mg/ml)	6.06
Drop composition (protein:well), microliters	2:10
Well volume (microliters)	1000
Sealant	grease
Seed	Repeat R5 #1
Constants in all wells	5.5% PEG 10K
Origin	R5 #1

2   Change in divalent metal cation vs. change in divalent metal concentration (+ microseed)

change in divalent cation

[Divalent cation] (mM) →

	200	300	400	500	600	700
$CaCl_2$	1	2	3	4	5	6
$CdCl_2$	7	8 X	9 X	10 X	11 X	12 X

Stock [SEB] (mg/ml)	6.06
Drop composition (microliter:microliter)	1.5:10
Well volume (microliters)	1000
Sealant	grease
Seed	R5 #1
Constants in all wells	5.5% PEG 10K
Origin	R5 #1
#of days before crystals appeared	5

Table 4b. Summary of refinement experiments for SEB: divalent cation series.(continued)

3   Change in [CdCl$_2$] (- microseed)

[Divalent cation] (mM) →

500	575	600	650	700	725
1 X	2 X	3 X	4 X	5 X	6 X
800	**875**	**950**	**1025**	**1100**	**1175**
7 X	8 X	9 X	10 X	11 X	12 X
1250	**1325**	**1400**			
13 X	14 X	15 X			

Stock [SEB] (mg/ml)	6.06
Drop composition (protein:well), microliters	2:02
Well volume (microliters)	1000
Sealant	grease
Constants in all wells	5.5% PEG 10K
#of days before crystals appeared	3
Origin	R5 #1

4   Change in divalent metal cation vs. change in divalent metal concentration (- microseed)

[Divalent cation] (mM) →

	2	3	4	5	6	7
CaCl$_2$	1	2	3	4	5	6
	20	**30**	**40**	**50**	**60**	**70**
CoCl$_2$	7	8	9	10	11	12
NiCl$_2$	13	14	15	16	17	18

change in divalent cation →

Stock [SEB] (mg/ml)	6.06
Drop composition (protein:well), microliters	4:02
Well volume (microliters)	1000
Sealant	grease
Constants in all wells	5.5% PEG 10K
Origin	R5 #1

Table 5. For BoNT A LC: the compositon of the successful reservoir solutions from the random screens

Random screen, reservoir solution #	R6 #16	R6 #18
Precipitant	2.0 M $(NH_4)_2SO_4$	20.6% PEG MME 5K
Buffer	0.1M Na acetate	0.1M Na succinate
pH	6.5	4.5
Co-precipitant		
Detergent	6.1% glycerol	
#of days before crystals appeared	32	32

Table 6. Summary of refinement experiments for BoNT A LC

1 Change in Volume% PEG MME 5K vs. change in buffer pH

%PEG MME 5K

change in pH	5%	10	15	20	25	30
4.5	1	2	3	4	5	6
5.5	7	8	9	10	11	12
7.5	13	14	15	16	17	18

Stock [BoNT A] (mg/ml)   5
Drop composition (protein:well), microliters   1.5:1.5
Well volume (microliters)   1000
Sealant   grease
Constant in all wells   0.1M Na succinate buffer
Origin   R6 #18

2 BoNT A complexed w/inhibitor: Inhibitor type vs. change in inhibitor concentration

change in inhibitor concentration in well (micromolar)

change in inhibitor type	130	260	390	520	650	780
P136	1	2	3	4	5	6
P190	7	8	9	10	11	12
control	13	14	15	16	17	18

Stock [BoNT A] (mg/ml)   5
Drop composition (protein:well:additive*), microliters   3:2.4:0.6
*additive   30% w/v xylitol
Well volume (microliters)   600
Sealant   grease
Constant in all wells   Hampton CS I #20:
25%PEG 4K, 0.1M Na acetate, pH 4.6, 0.2M $(NH_4)_2SO_4$

References

Bavari, S., Ulrich, R.G. and Dyas, B. (1996) *J. Infect. Dis.*, **174**: 338

Jardetzky, T.S., Brown, J.H., Gorga, J.C., Stern, L.J., Urban, R.G., Chi, -in, Stauffacher, C., Strominger, J.L. and Wiley, D.C. (1994) Three-dimensional structure of a human class II histocompatibility molecule complexed with superantigen. *Nature*, **368**: 711

Lacy, D.B., Tepp, W., Cohen, A.C., DasGupta, B. R. and Stevens, R.C. (1998) Crystal structure of botulinum neurotoxin type A and implications for toxicity. *Nature Structural Biology*, **5** (10): 898

Papageorgiou, A.C., Tranter, H.S. and Acharya, K. R. (1998) Crystal structure of microbial superantigen Staphylococcal enterotoxin B at 1.5 Angstroms resolution: implications for superantigen recognition by MHC class II molecules and T-cell receptors. *J. Mol. Biol.*, **277**: 61

Segelke, B., Ringhofer, S. and Rupp, B.
http://www-structure.llnl.gov/crystool/crystool.htm

Trakhanov, S. and Quiocho, F. A. (1995) Influence of divalent cations in protein crystallization. *Protein Science*, **4**: 1914

Trakhanov, S., Kreimer, D. I., Parkin, S., Ferro-Luzzi Ames, G. and Rupp, B. (1998) Cadmium-induced crystallization of proteins: II. Crystallization of the *Salmonella typhimurium* histidine-binding protein in complex with L-histidine, L-arginine, or L-lysine. *Protein Science*, **7**: 600

Ulrich, R.G., Bavari, S. and Olson, M.A. (1995) *Nature Structural Biology*, **2**: 554

(1999) Protein Crystallization: Techniques, Strategies, and Tips. T.M. Bergfors, ed. International University Line: La Jolla, CA, USA

Acknowledgements

Comparison of a Stellar Evolution Code to Helioseismology*

Davienne N. Monbleau

Rensselaer Polytechnic Institute

Mentors: Dave Dearborn and Stephen Murray

Lawrence Livermore National Laboratory

Livermore, California 94551

December, 17, 1999

Prepared in partial fulfillment of the requirements of the Undergraduate Research Semester under the direction of Dave Dearborn, Research Mentor, in the Lawrence Livermore National Laboratory.

ABSTRACT

Stars provide the basic metric for setting the age and size of the universe. Their nuclear yields drive the chemical evolution of our galaxy and others. Thus, understanding the evolution of stars is essential to understanding the evolution of the universe. However, about half of all stars are binaries, many of them deformed by their rapid rotation. These non-spherical systems have many exotic behaviors, including Type I supernovae, cataclysmic variables, and novae. Without a more complete understanding of these objects, our knowledge of the evolution of the universe will be limited.

Thus, our ultimate objective is to develop a three-dimensional stellar evolution code, which will emulate the thermodynamic configuration of non-spherical stellar systems and their evolution through time. This is known as Djehuty, A Next Generation Stellar Evolution Code[1]. In preparation, we are testing the accuracy of the basic input phyiscs of our one-dimensional code by comparing sound-speed verse depth from an evolved solar model to observed helioseismic data. The code may then be applied to other stars to provide more precise evolutionary models and further the development of a three dimensional code.

1. INTRODUCTION

Stellar evolution describes the lifetime of a star. It links different stellar types as distinct phases in the lifetimes of individual stars. The evolution is primarily driven by balancing radiative energy loss with gravitational or nuclear energy. The fusion reactions inside the core slowly alter the star's interior composition and the star's structure evolves. Depending on the star's initial mass and composition, it may go through numerous burning phases from Hydrogen to Silicon. After millions of years, very massive stars will be left with an Iron core.

There are five primary equations in our stellar evolution code:

1. Hydrostatic equilibrium

2. Energy transport

3. Thermal equilibrium

4. Mass continuity

5. Thermodynamic equilibrium

Hydrostatic equilibrium is the balance between the inward gravitational pressure and the outward radiative pressure. Energy transport accounts for the convective energy between the zones, which alters the luminosity at each zone. Thermal equilibrium is the conservation of energy. This can be represented by either the diffusion of energy by radiation or the transport of energy by convection. Thermodynamic equilibrium is the equation of state. It relates the pressure, temperature, density, and composition of the star. Mass continuity states that the sum of masses in each zone equals the total mass of the star. Other important quantities that need to be considered are the nuclear reactions rates, neutrino emission rates, neutrino loss rates, and the opacities[2] of the regions within the star.

Combining all of the above with an initial mass and composition, we can predict the pressure, temperature, luminosity, and radius of the star at any point in time.

2. PURPOSE

The development of a three-dimensional stellar evolution code is a three year project. However, we can not proceed further until we are satisfied with the accuracy of the physics models in our current one-dimensional code. Precise basic input physics is essential. Thus, we are using helioseismic data as a metric for the code's precision, particularly the equation of state.

Helioseismolgy is a form of acoustical spectroscopy. It uses observed frequencies of solar oscillation to infer properties of the solar interior with an accuarcy of 0.1%, such as the speed of sound.

3. ANALYSIS

We modified the code to fit our Sun. Unlike stars, which are too far away, the Sun is close enough to do other research, such as helioseismology. As a result, we have more accurate data about the Sun, so we increased the number of zones from 300 to 500. Then we evolved a $1M_\odot$ star 4.55×10^9 yrs, the age of our Sun[3], to see if an accurate model was generated. The values we are using as a basis for the Sun are:

1. $L_\odot = 3.845 \times 10^{33}$ ergs

2. $R_\odot = 0.69595 \times 10^{11}$ cm

3. $M_\odot = 1.98892 \times 10^{33}$ g

4. $t_\odot = 4.55 \times 10^9$ yrs

However, the final model did not have the correct radius and luminosity values. That we altered the mixing length, α, and the hydrogen composition, x, until we had $1R_\odot$ and $1L_\odot$. The mixing length was 1.705 and the hydrogen mass fraction was 0.69075. The radius and luminosity where within 4 and 3 decimal places, respectively. Figure 1 shows the HR diagram of the Sun, derived from the stellar evolution code.

The helioseismic sound-speed data that we used as a comparison is from Basu, S. et al.(1997)[4]. This data was taken from Model S in Christensen-Dalsgaard, J. et al. (1996)[5]. Figure 2 shows the inversion results for the sound-speed verse stellar radius. Figure 3 shows our calculated sound-speed verse stellar radius.

Figure 4 shows the residuals of the two previous plots. There is an average accuarcy of about 1%. However, we would like to acheive an accuracy of about 0.1%.

4. CONCLUSIONS

Although the current 1-dimensional code is more accurate than we had anticipated, we would like to increase that accuracy tenfold. In order to do this we need to develop a more precise equation of state. We are currently modifying the code to calculate more accurate pressures within each zone. It will use the OPAL equation of state tables[6] to interpolate a correct chemical potential and hence, a more accurate pressure.

REFERENCES

[1] http://ldrd.llnl.gov/cgi-bin/propupdate

[2] Igleseas, C. A., Rogers, F. J. 1996, ApJ, 464, 943

[3] Bahcall, J. N., Pinsonneault, M. H., Wasserburg, G. J. 1995, *Rev. Mod. Phys.*, 67, 781

[4] Basu, S., Chaplin, W. J., Christensen-Dalsgaard, J., Elsworth, Y., Isaak, G. R., New, R., Schou, J., Thompson, M. J., Tomczyk, S. 1997, MNRAS, 291, 243

[5] Christensen-Dalsgaard, J. et al. 1996, *Science*, 272, 1286

[6] Rogers, Forrest J., Swenson, Fritz J., Iglesias, Carlos A. 1996, ApJ, 456, 902

This preprint was prepared with the AAS LATEX macros v4.0.

Solar HR Diagram
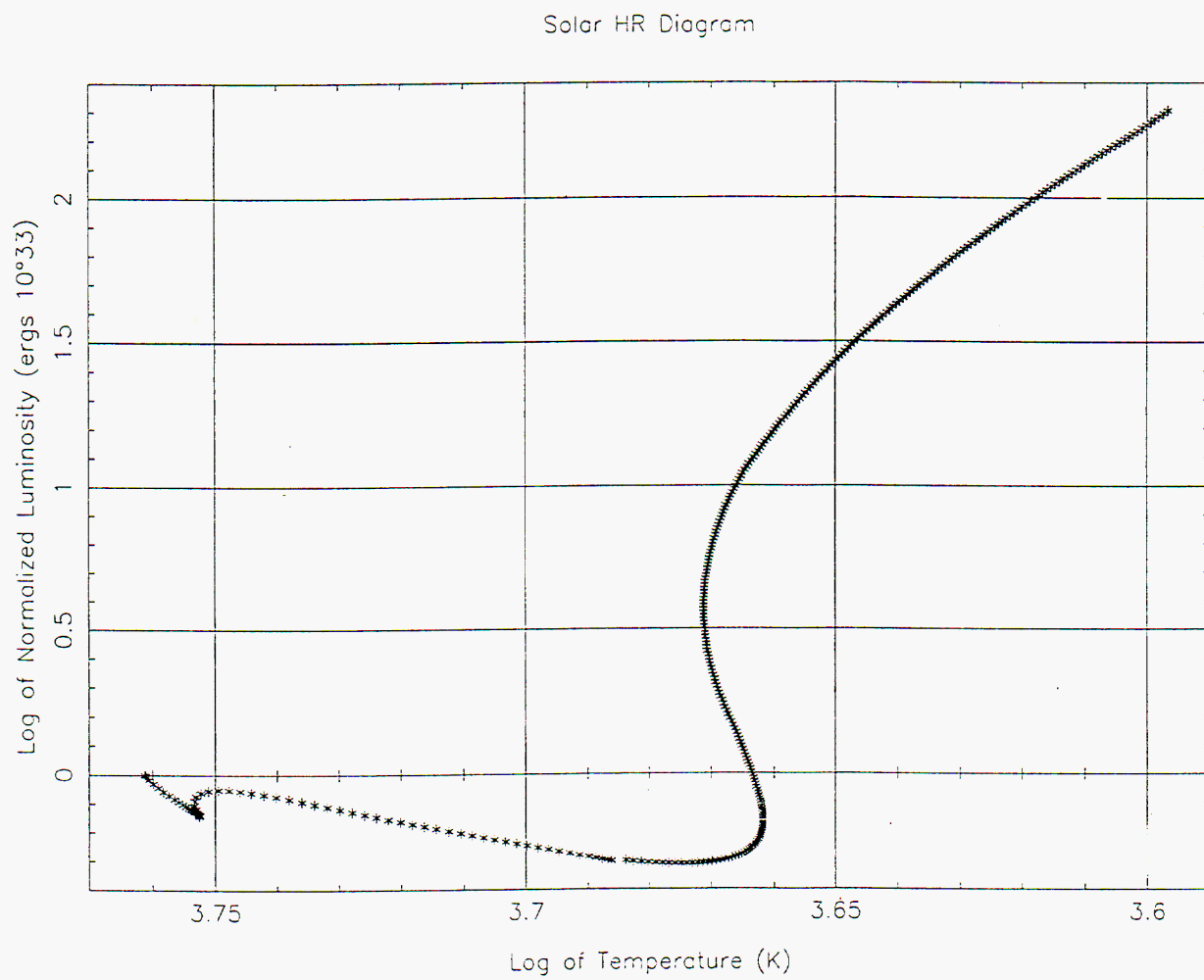
Fig. 1.— Calculated solar evolution from Stellar Evolution code.

Helioseismology: Calculated Sound Speed vs. Stellar Radius
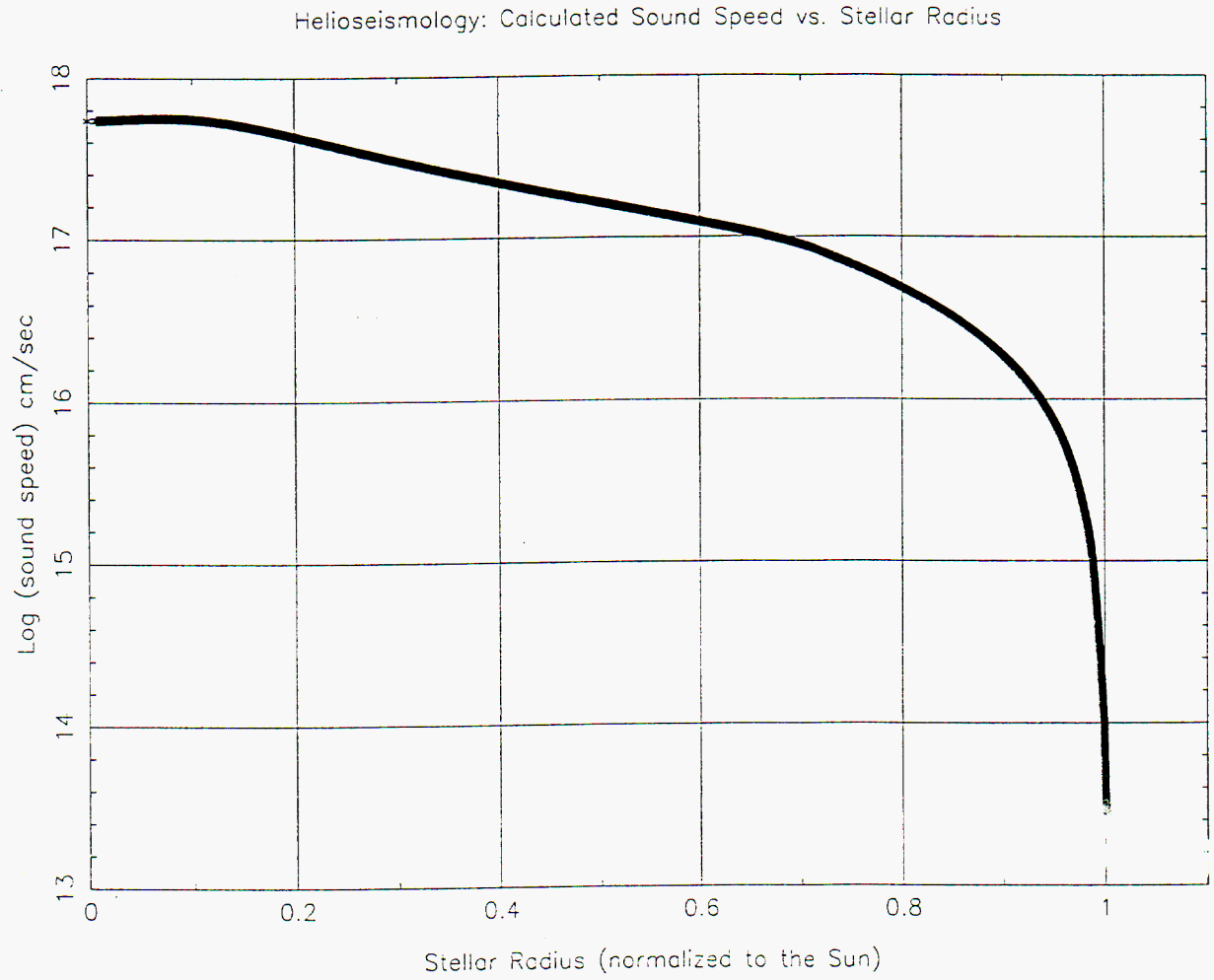
Fig. 2.— Calculated from the helioseismic inversion results from Basu et al. (1997).
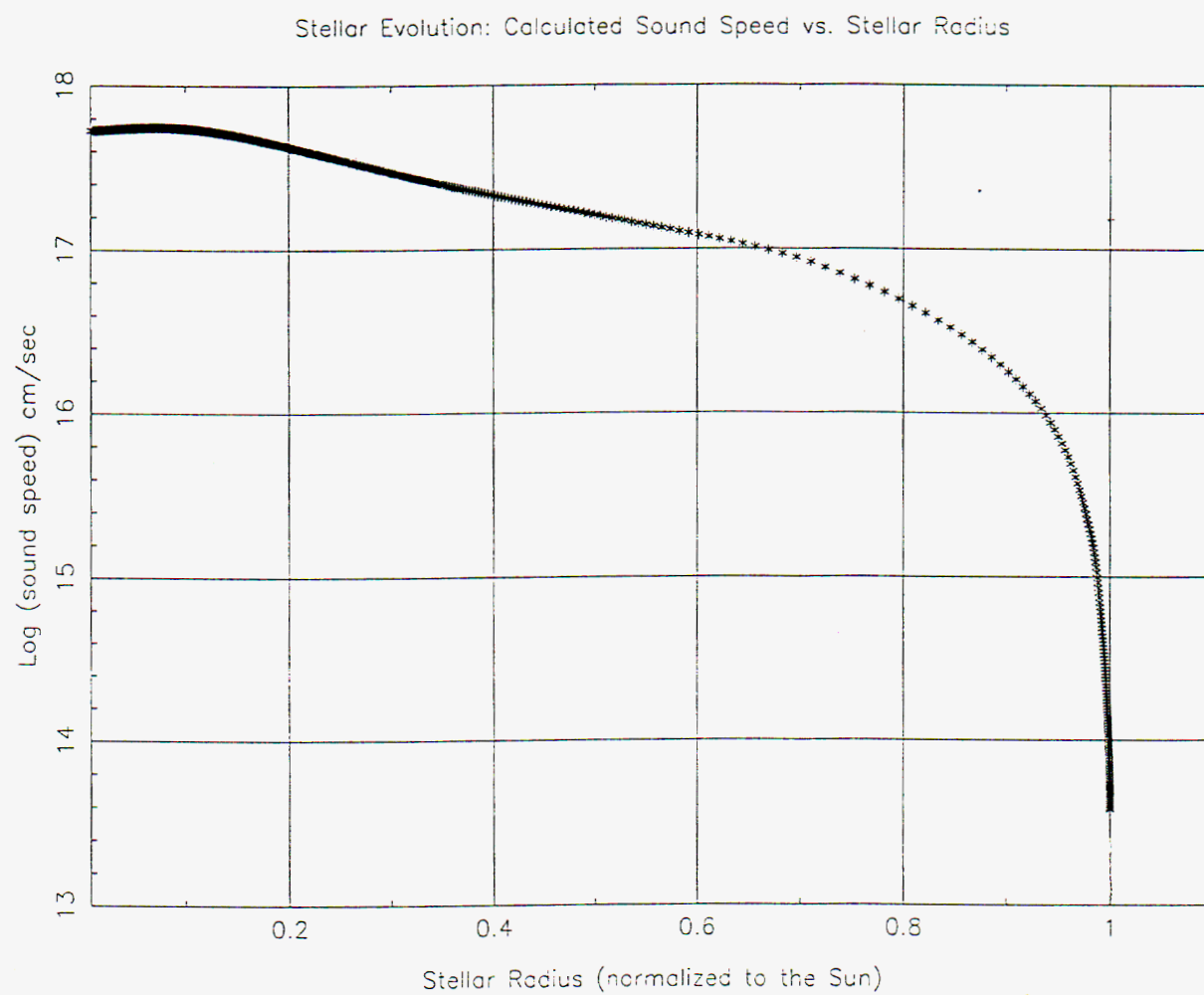
Fig. 3.— Calculated from our stellar evolution code.
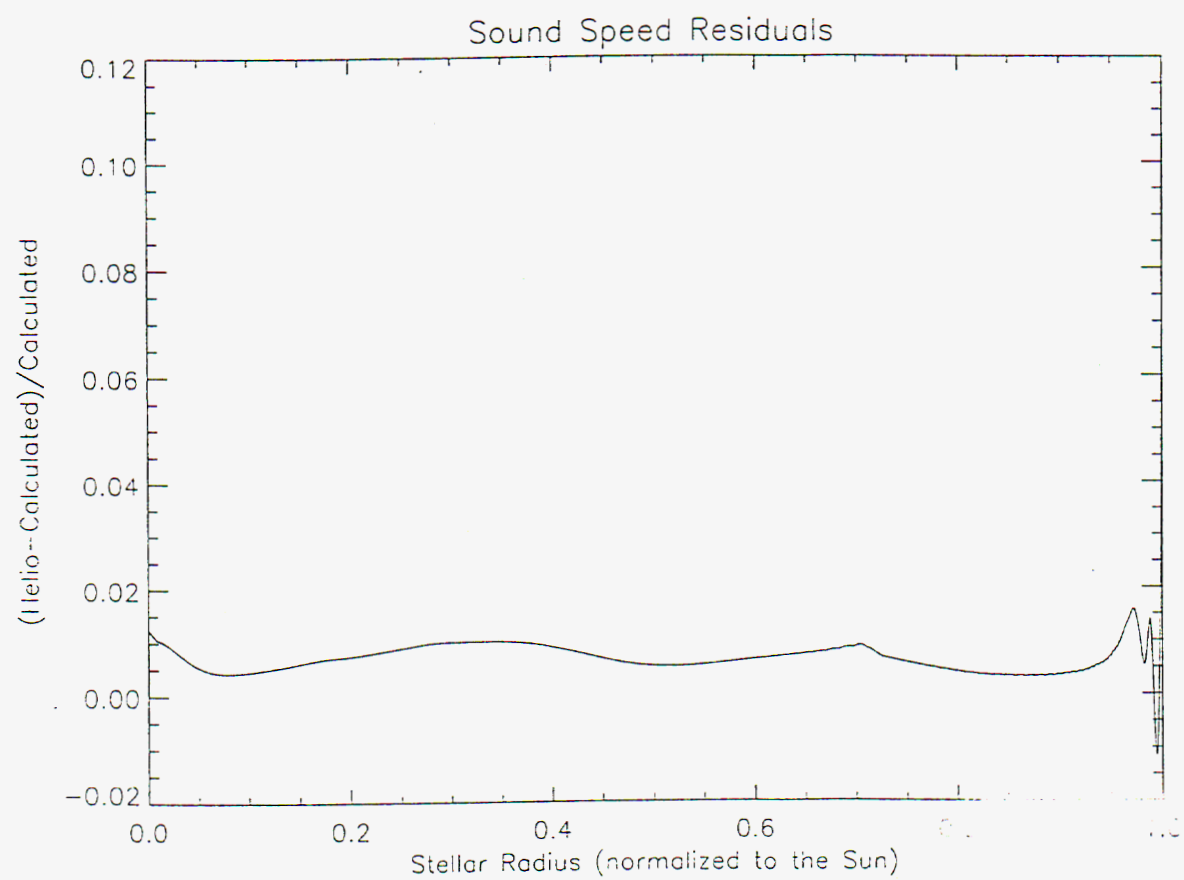
Fig. 4.— Sound speed residuals.

Parallel Object-Oriented File I/O
for the Sapphire Project*

David Nault

University of Cincinnati

Lawrence Livermore National Laboratory
Livermore, California 94551

December 10th, 1999

Parallel Object-Oriented File I/O
For the Sapphire Project

David Nault, University of Cincinnati
Center for Applied Scientific Computing

DESCRIPTIVE ABSTRACT

The purpose of the Sapphire Project is to provide researchers
with a set of tools for interactive exploration of large, complex,
multi-dimensional data sets. In order to solve this problem we
are applying and extending ideas from data mining and pattern
recognition in our development of a scalable, parallel
application toolkit in C++. Various aspects of the application
toolkit will be discussed, with the primary focus on the file
input/output system and the parallel communication
infrastructure.

1 Introduction

The Sapphire Project is a software development effort underway at the Center for Applied Scientific Computing at Lawrence Livermore National Laboratory. Sapphire is an object-oriented toolkit for the interactive exploration of large, complex, multi-dimensional scientific data. We are extending and applying ideas from data mining and pattern recognition to design and implement a new generation of tools to help scientists extract useful information from data.

Data mining begins with the extraction of key features from the raw data. This may involve reducing the dimensionality of the data set and the application of noise reduction, edge detection, and various other algorithms. Wavelet functions are another useful way to characterize the features of the data.

These features are then fed into a decision tree for classification and pattern recognition. The type of data and the methods used to process it will vary depending on the needs of our clients.

This paper will discuss the development of the Sapphire toolkit's file input/output subsystem, which will be able to read data files in several popular formats and translate them into the internal representation used by the Sapphire feature extraction routines.

1.1 Development Environment

The Sapphire toolkit is being developed using C++ on Sun workstations. The Message Passing Interface (MPI) standard is being used for inter-process communication. One of the benefits of this arrangement is that MPI allows a single workstation to emulate a parallel computer with any number of processors. This means that code can be developed, tested, and debugged on the Suns before it is run on the real parallel machines.

Compatibility across a various UNIX platforms is a priority. Most if not all of the code will be written in ANSI C++. Features of the language that are not widely supported by modern compilers will not be used.

1.2 Third Party Libraries

When we read files, we would like to leverage third party code libraries that have been developed for this purpose. For example, when reading Flexible Image Transport System (FITS) files we can use the FITSio library to simplify the process.

A large part of the work that goes into Sapphire file I/O lies in elegantly incorporating these libraries into the toolkit and adapting them to work efficiently on parallel computers.

2 Important Classes

One of the benefits of programming in C++ is that functional components can be modeled as objects. We created a handful of C++ classes to handle the responsibility of reading and translating data files.

There are three key classes that play a role in reading a file:

1) RegData – represents the internal file format

2) Boxes – store information about rectangular regions

3) FITSFile – a wrapper class for FITSio.

2.1 RegData

The Sapphire data format is called RegData (for Regular Data). The primary responsibility of the file I/O system is to convert data files into this format.

We designed RegData to allow for direct access to data. The core of the RegData class is an array of data points arranged in a uniform fashion. No iterators or advanced accesors are provided. A pointer to the base address of the array is the single entry point. This means that clients of RegData must be aware of the ordering of the data in the array. The benefit of this approach is that clients can loops over the data directly, eliminating the overhead of repeated calls to accessor functions.

2.2 Boxes

The data we are working with is rectangular, so naturally we need some way to think about the boundaries of all or potions of the data. Boxes are used to delimit regions of the index space and to perform calculations on rectangular regions. The Box class has functions for manipulating and intersecting these regions.

2.3 FITSFile

FITSFile is a wrapper class that makes calls to FITSio, a third party utility library for reading and writing FITS files. The wrapper class provides a uniform interface for working with FITS files, and insulates the client code from the details of FITSio, which in turn conceals the messy details of the FITS format.

FITSFile is a subclass of DomainData, an abstract base class that provides a uniform interface for working with files in *any* format. When we want to add support for a new format, we will create a new subclass of DomainData and flesh it out with code to perform the required translation.

3 Two Strategies for Parallelism

One of the responsibilities of the file I/O system is to make sure that as data files are read in portions of them are distributed properly across multiple processes. We have explored two possible ways of doing this.

3.1 File Server Strategy

When using this strategy, one process is designated the file server. The file server is solely responsible for reading the entire file and converting it to RegData. This RegData object is then cut up into several pieces that are sent to the other processes.

When data needs to be written back to disk, the file server waits for the other processes to send it their pieces of the RegData. The server stitches the RegData back together and writes it to disk.

This strategy is appealing because the translation from the raw data into RegData is done all at once. Consequently, the code for performing the translation does not have to deal with special cases (like the data you are reading not being contiguous in the file).

One of the drawbacks of this method is that after the server has read the file it must send all the pieces of it across the network to the other processes. This may be an extremely time consuming operation depending on the size of the file and the speed of the network.

3.2 Direct Access Strategy

The Direct Access Strategy cuts out the middleman; every process is responsible for handling its own file I/O. Each process determines which part of the file it should read from the disk and then converts that small portion to RegData.

Writing to a file in this manner is a bit trickier. When several processes a
to a file at the same time, there is the potential for the requests to overlap and for the file to become corrupted. We need to ensure that only one process is writing to a file at a time. One way to do this is to pass a token that say, "Okay it's your turn to write." The logic for the token passing is straightforward:

```
If you are the root process, write the first part of the file and send
the token to the next process.

Otherwise wait for someone to send you a token, write your portion of
the file, and pass the token along (unless you are the last process).
```

Although conceptually simple, the direct access method requires the file translation code to be able to perform partial reads, which may be a non-trivial operation, especially when the target data is not contiguous. Direct access may, however, prove to be more efficient because it does not require that any data be sent between processes.

4 Results

We created prototype applications for both of the file I/O strategies. Initial results on the Sun workstations indicate that the Direct Access strategy is markedly more efficient than the File Server strategy, probably due to the low bandwidth available for inter-process communication on these machines. It remains to be seen whether this will be a serious issue when the code is transferred to the high-end parallel computers.

As the Sapphire code base grows and support for more file formats is needed, the File Server strategy may prove to be easier to maintain and extend. The file translation code for this strategy is much simpler, so it will be easier to write translators for new file formats.

In addition to these prototypes, our work with the FITS file translator led to the development of a GUI application for the X Window system that allows a user to display a two-dimensional FITS image. This viewer will be a useful diagnostic tool for interpreting the results of the image processing functions in the Sapphire toolkit.

5 Summary

Sapphire is an object-oriented toolkit currently under development at the Center for Applied Scientific Computing. It will enable scientists to extract useful information from massive data sets using data mining and pattern recognition techniques.

Before the data from the scientists can be analyzed, it must be read and converted into a universally recognizable format. The part of the toolkit that handles this conversion is also responsible for ensuring that the translated data is distributed properly across multiple processes.

We have discussed some of the C++ classes that perform the file translation work, and have outlined two competing strategies for performing parallel file input/output.

Now that Sapphire can read and recognize data from FITS files, the foundation is in place for the development of the image processing and feature extraction components of the toolkit.

Removing Nitrate and Perchlorate from Groundwater with Anaerobic Microbial Films*

Mai Ho
Georgia Institute of Technology
Nathaniel Nicks
University of Washington

Lawrence Livermore National Laboratory
Livermore, California 94551

12/10/99

Table of Contents

Removing Nitrate and Perchlorate from Groundwater with Anaerobic Microbial Films

Mai Ho
Georgia Institute of Technology
Nathaniel Nicks
University of Washington

Environmental Restoration Division

Descriptive Abstract

The purpose of this project was to design a remediation unit that will lower nitrate and perchlorate concentrations in the groundwater at a DOE material test site (Site 300). The design process included evaluation of an existing bioreactor and consulting with lab engineers, researchers, and technicians to build a more versatile and improved unit. The bioreactor will be a secondary treatment facility following a conventional solvent removal unit. The design consists of two packed towers operated in series. Microorganisms will be added to the bioreactor to colonize and form a biofilm. The addition of ethanol to the reactor will allow the biofilm to metabolize the nitrate and perchlorate contaminants into harmless components: nitrogen gas, chloride ion, and water.

Date   Typed:12/07/99
Revision #2

Background

The bioreactor will be installed at Site 300, a DOE test site remotely located in the Altamont Hills. High levels of volatile organic compounds (VOCs), nitrate, and perchlorate have been detected in the groundwater. Maximum contaminant levels (MCLs) for nitrate and perchlorate have been set at 45,000 µg/L (45 mg/L) and 18 µg/L, respectively. The bioreactor is a secondary remediation unit that will follow a conventional VOC treatment unit. The bioreactor should reduce nitrate and perchlorate concentrations to meet the regulatory limits.
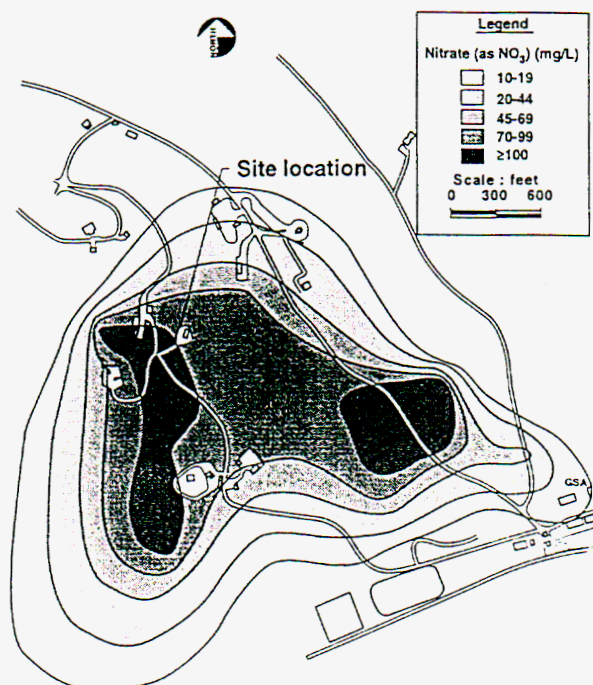
Figure 1: Site map showing nitrate concentrations in groundwater in the vicinity of the proposed bioreactor location. In the high concentration zones of the contours, nitrate levels exceed the MCL of 45 mg/L.
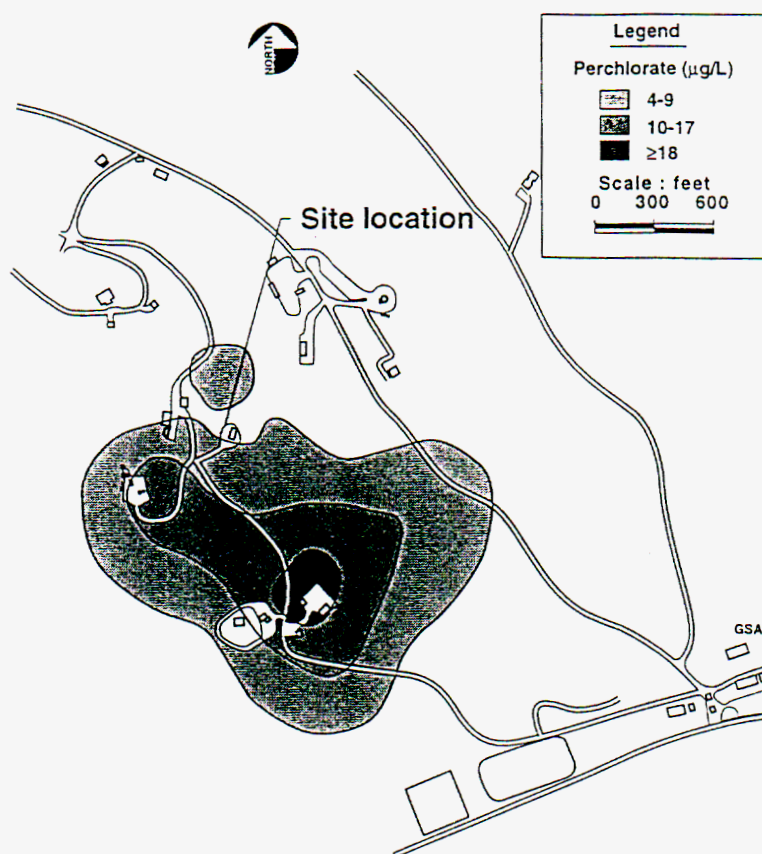
Figure 2: Site map showing perchlorate concentrations in groundwater in the vicinity of the proposed bioreactor location. In the high concentration zones of the contours, perchlorate levels exceed the MCL of 18 µg/L.

Objectives

Design a bioreactor that can:

A) reduce groundwater nitrate concentration from 90 mg/L to 20 mg/L (>75% removal).

B) reduce groundwater perchlorate concentration.

C) be operated at flow rates of up to 8gpm.

Strategy

The proposed bioreactor makes use of facultative anaerobic microorganisms to reduce and destroy inorganic pollutants (nitrate and perchlorate). Microorganisms will settle on

the inner surfaces of the reactor (packing material) and form a mature biofilm over time when a carbon source in the form of ethanol is provided. The upward-flow bioreactor is designed to optimize the degradative process. Ethanol is metered into the groundwater treatment stream to facilitate the transformation of nitrate (NO_3^-) and perchlorate (ClO_4^-) to benign compounds: nitrogen gas, water, chloride ion, and carbon dioxide derived from ethanol.

Bioreactor   **Packing material**   **Biofilm**

$$N_2 + Cl^- + CO_2 + H_2O$$

$$NO_3^- + ClO_4^- + C_2H_5OH$$

Plastic packing
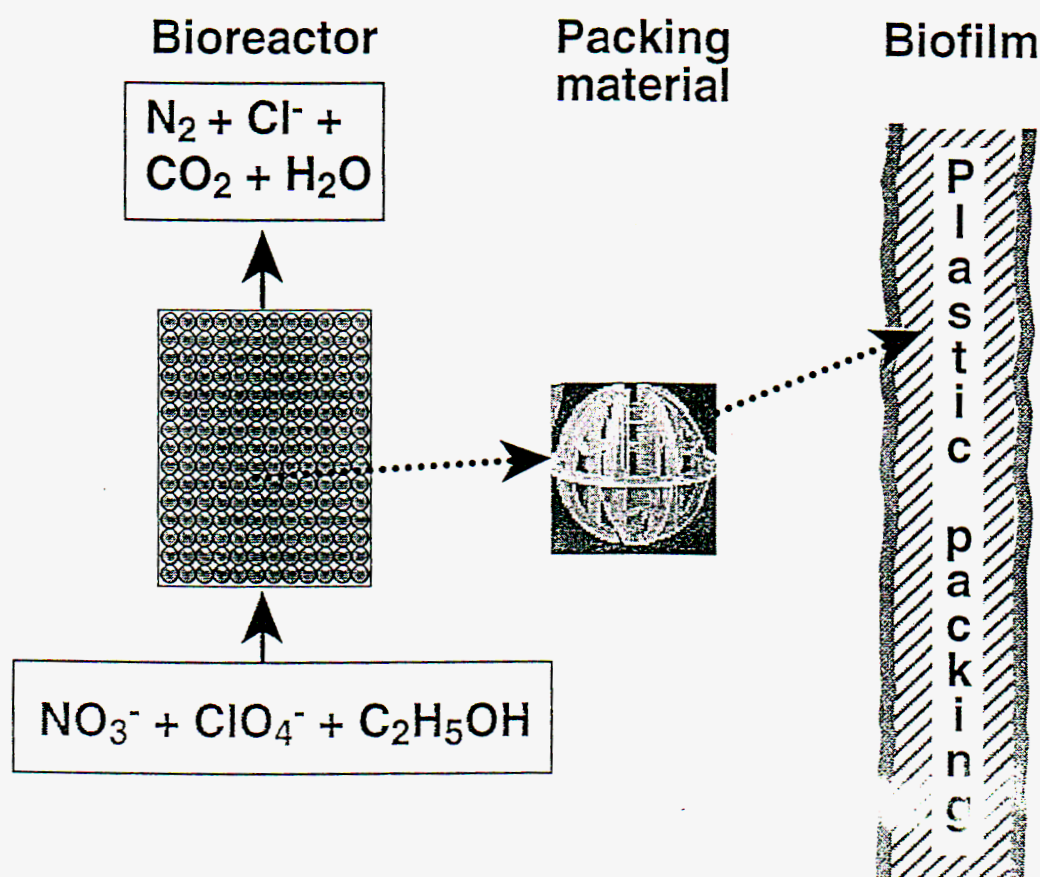
Figure 3: NO_3^-, ClO_4^-, and C_2H_5OH (ethanol) are metabolized into N_2 (gas), Cl^- (ion), CO_2 (gas), and H_2O (water) by the biofilm.

Results

The initial step in the design process was to determine the kinetic parameters of denitrification. The analysis of experimental data (see Appendix A) from an existing bioreactor (BTU-1) suggested that nitrate removal followed first-order kinetics. This mathematical relationship was used for modeling the new bioreactor (detailed kinetic explanations/calculations are presented in Appendix B).
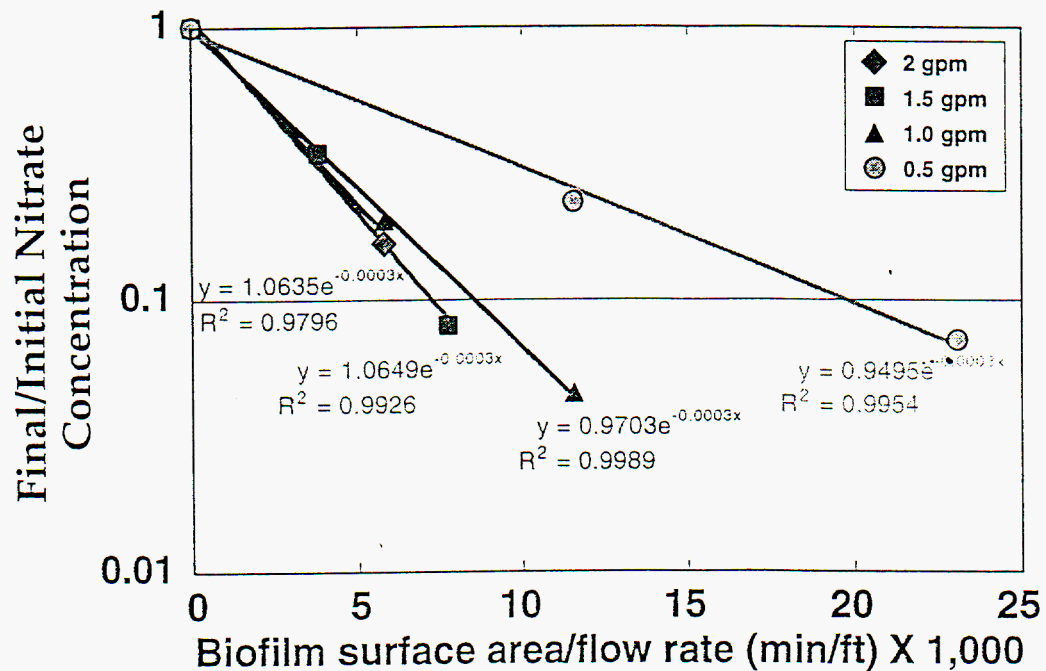
Figure 4: Input data from BTU-1 demonstrating first order kinetics. The effective removal efficiency is plotted versus a normalized bioreactor surface area calculation.

The ratio of the final to initial nitrate concentration was plotted on a semi-log scale versus bioreactor surface area normalized for treatment flow rate (as seen in **Figure 4**). The nitrate removal efficiency, which is inversely proportional to the final/initial nitrate concentration ratio, increases with normalized surface area. Theoretically, the linear relationships for all flow rates on the log plot should follow the same trend; however, a deviation occurs at the 0.5 gpm flow rate resulting in a flatter slope or a decrease in the overall removal efficiency. This deviation is probably due to poor mixing in the

bioreactor, which would decrease the effective surface area of the biofilm and lower the reactor removal efficiency.

An additional BTU-1 batch experiment was conducted to compare nitrate and perchlorate degradation as a function of time. It aided in predicting perchlorate degradation of the new reactor (see Appendix C).

The groundwater at the treatment location is significantly colder than the water used during the BTU-1 experiments. This necessitated correction of the degradation (see Appendix D).

Predictions for the bioreactor's contaminant removal capabilities are presented in graphical format. Contaminant removal efficiency is dependent on biofilm surface area (see **Figure 5a**) and treatment flow (see **Figure 5b**).
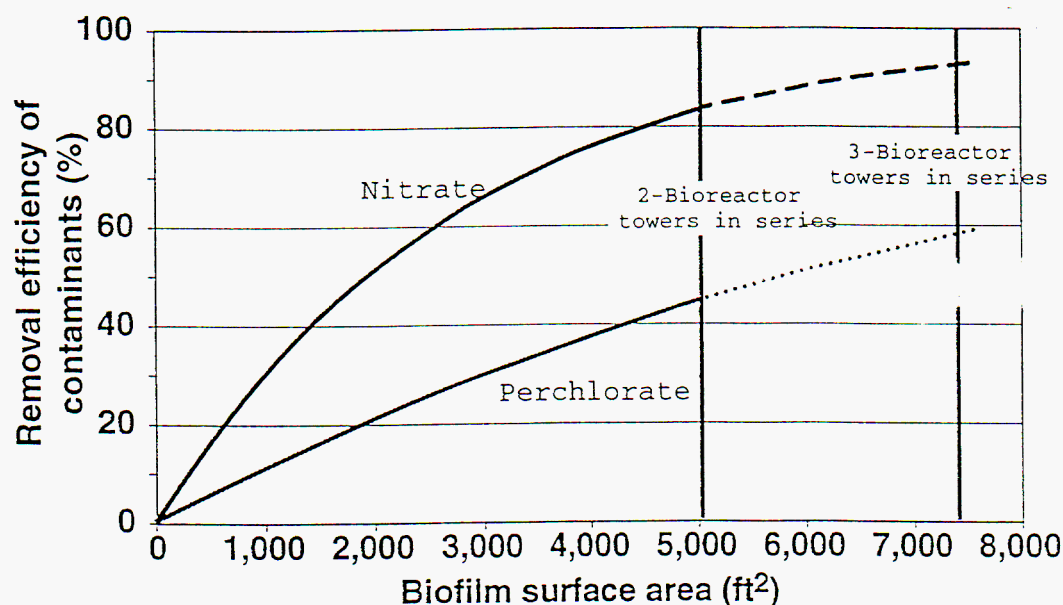
Figure 5a: Predicted contaminant removal efficiencies increase with biofilm surface area. Dashed line extensions indicate performance of a third bio-tower.
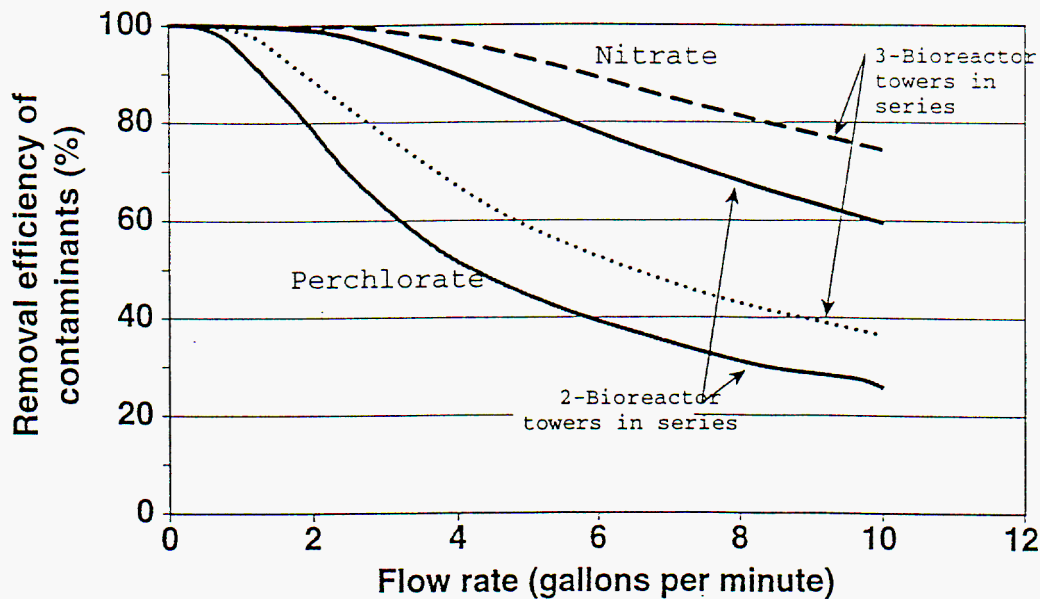
Figure 5b: Predicted removal efficiencies decrease with increasing treatment flow rate. Dashed lines indicate performance of a third bio-tower.

Predicted removal efficiencies for both nitrate and perchlorate increase with biofilm surface area and decrease with increasing treatment flow. The removal efficiency for nitrate is greater than that of perchlorate according to the estimated degradation rate constants (see Appendix C).

In order to maximize the contaminant removal efficiencies, a packing material with a high surface-area-to-volume ratio should be used. Maximizing the amount of surface area in a given volume is desirable for providing plenty of room for bacterial deposition while keeping the treatment unit compact. As shown in **Figure 5b**, an increase in treatment flow decreases the removal efficiencies because the time available for microorganisms to degrade the contaminants is reduced. The predicted bioreactor performance data suggest that at a 5 gpm treatment flow, 2 bioreactor towers operated in series would be sufficient for removing 80% of the nitrate and 45% of the perchlorate.

The addition of a third tower would yield similar removal efficiencies at an even higher treatment flow rate (8 gpm).

The bioreactor tower was designed for easy maintainability (see **Figure 6**).
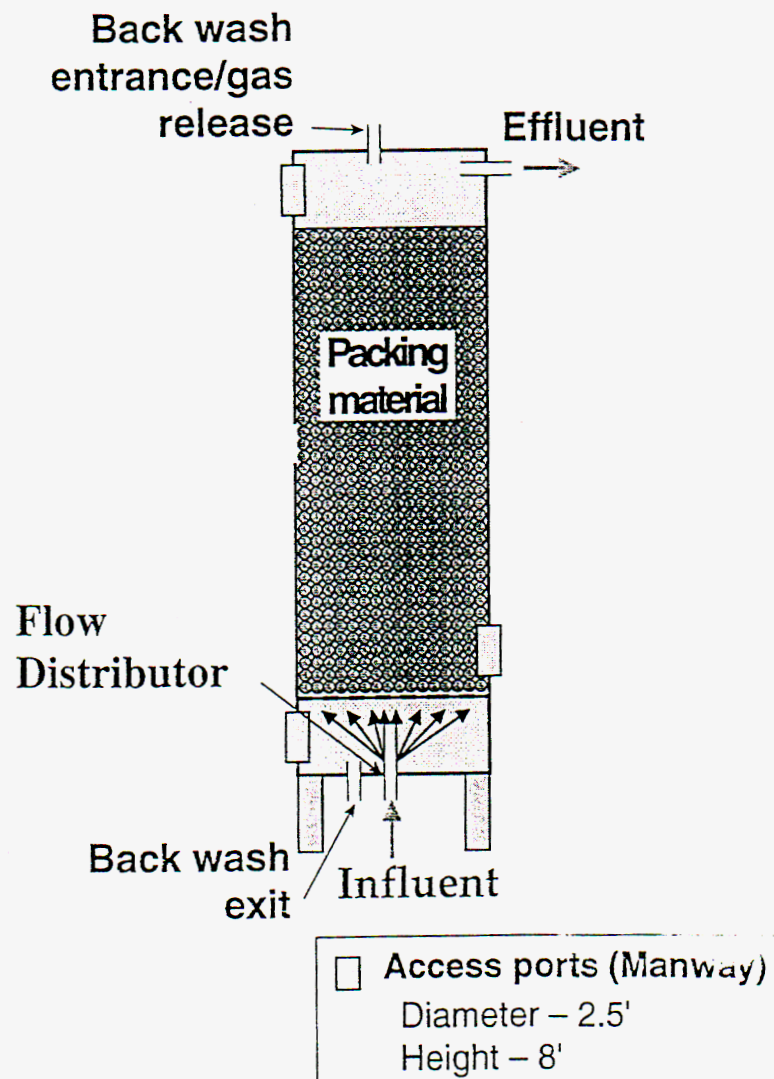
Figure 6: Diagram of the upward-flow bio-tower.

The bio-tower is filled with packing material that provides a surface for the microbes to grow on and form an anaerobic biofilm. During normal operation, the contaminated water will flow upward through the tower, past the biofilm-covered packing material and

exit out the top. A flow distributor is provided at the intake to ensure even flow distribution upon entry. Backwashing ports will operate in a counter-flow configuration to remove excessive biomass. Access ports at the top and bottom of the tower allow for convenient packing addition and removal, respectively. An additional access port is placed at the flow distributor to allow for repairs or possible replacement.

The piping schematic of our bioreactor indicates the different paths of treatment flow (see **Figure 7**).
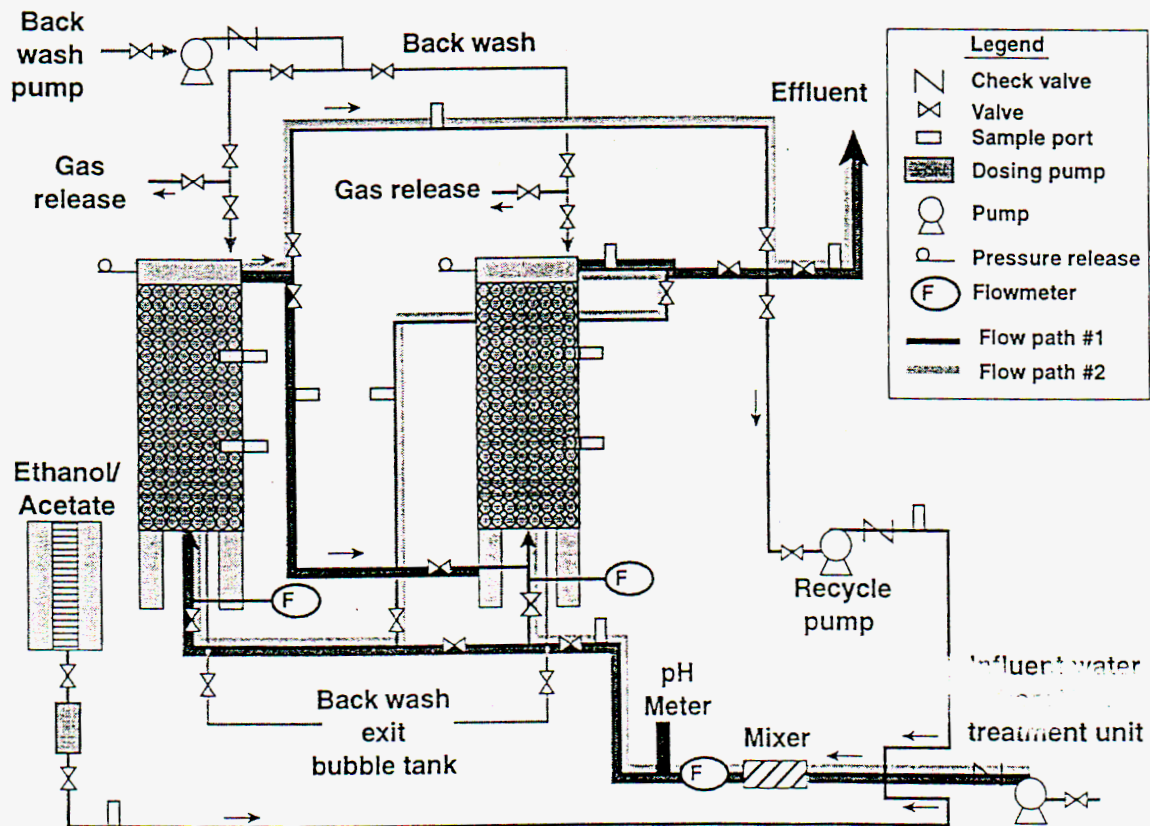
Figure 7: Diagram of the proposed 2-stage bio-treatment unit.

In flow path # 1 (red line) the contaminated water enters the bottom of the tower on the left, exits out the top and flows through the bottom of the tower on the right and exits out the top. In flow path # 2 (blue line) the flow sequence is reversed and the first reactor becomes the second in the treatment flow path. Periodically changing the flow path will

volume ratio is desirable for achieving a compact treatment unit, where as its large porosity minimizes pressure drop across the reactor, thereby lowering strain on the pumps.

Ethanol is an effective carbon source (electron donor) for biological denitrification (Kappelhof, 1992). For safety purposes, the ethanol will be diluted in water to lower the vapor pressure of the aqueous mixture to below the flammable limit (see Appendix F).

Conclusions

The use of anaerobic fixed-film reactors is a simple, efficient and cost-effective technology for removing nitrate and perchlorate from groundwater. Positive attributes of the new reactor presented in this report include its compactness, its transportability, and the reduced maintenance requirements when compared to fluidized-bed reactors. Most importantly, it is superior to contaminant-separation technologies (e.g., ion exchange processes) due to its ability to destroy contaminants on-site without generating secondary, toxic waste.

Acknowledgements

References

Fogler, Scott H. <u>Chemical Reaction Engineering</u>. 2nd edition. Prentice Hall PTR, 1992. p. 50.

Kappelhof, J.W.N.M., van der Hoek, J.P. and Hijnen, W.A.M. <u>Experiences with fixed-bed denitrification using ethanol as substrate for nitrate removal from groundwater</u>. Water Supply, 1992. pp. 91-100.

Parker, A.L., Sikora, L.J., Hughs, R.R. <u>Denitrification Kinetics in Packed Beds</u>. AlChE Journal, Vol.22 No. 5, September 1976. pp. 851.

Perry, Robert H., Green, Don W. <u>Perry's Chemical Engineers' Handbook</u>. 7th edition. McGraw Hill, 1997. pp. 2-235.

<u>Plastic Jaeger Tri-Packs</u> URL: www.jaeger.com

Appendix A
Experimental data of BTU-1

Nitrate Test Data

Test1, Flow rate = 0.5 gpm, = 1.9 Lpm

Sample	Time (hours)	Chemical Oxygen Demand (mg/L)				Ethanol Concentration (mg/L)				Nitrate Concentration (mg/L)				% Removal Reactor A	% Removal Reactor B	% Removal Total
		PTU	RAI	RAO	RBO	PTU	RAI	RAO	RBO	PTU	RAI	RAO	RBO			
S0	0	0	60	11	0	0.0	34.2	8.3	0.0	26.24	99.84	14.88	4.66	85.1	10.3	95.3
S1	1										98.03	14.48	5.46	85.2	9.2	94.4
S2	2		50	7	2		28.9	6.2	3.5		94.51	23.87	5.39	74.7	19.6	94.3
S3	3										95.52	28.42	5.77	72.3	21.6	94.0
S4	4		65	6	6		36.9	5.7	5.7		97.19	26.34	7.96	72.9	18.9	91.8
S5	5										101.15	25.43	8.97	74.9	16.3	91.1
S6	6		65	13	0		38.9	9.4	0.0		103.26	22.53	8.80	78.2	13.3	91.5
S7	7										103.68	25.5	8.43	75.4	16.5	91.9
S8	8		60	23	0		39.0	14.7	0.0		106.52	25.47	8.24	76.1	16.2	92.3
Average Values		0.0	61.8	12.0	1.6	0.0	35.2	8.8	1.8	26.2	99.9	22.8	7.1	77.2	15.8	93.0

Nitrate Removal, Flow = 0.5 gpm

Nitrate Concentrations, Flow = 0.5 gpm

Ethanol Concentrations, Flow = 0.5 gpm
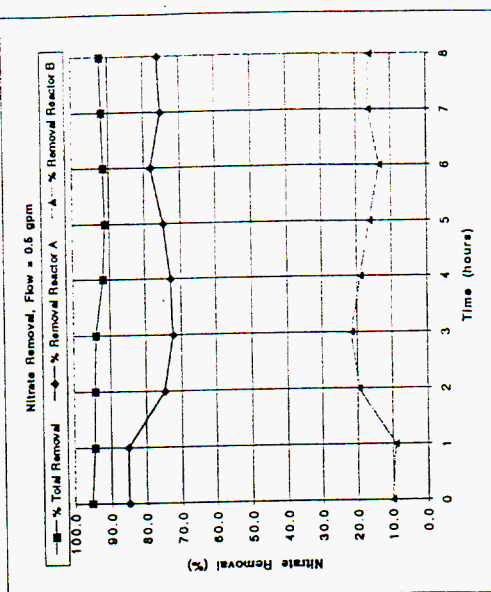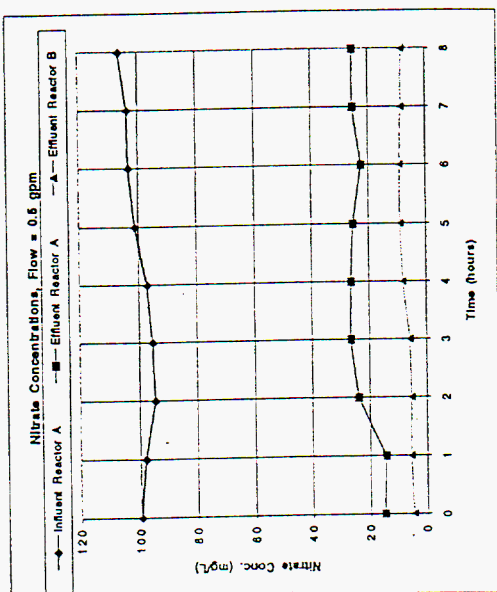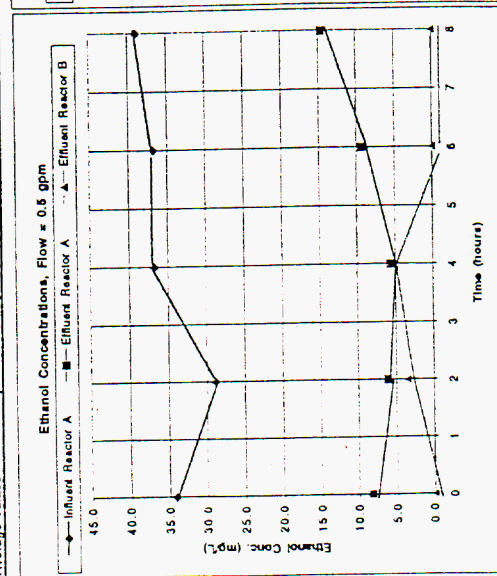
Nitrate Test Data

Test2, Flow rate = 1.0 gpm = 3.8 lpm

*results too high, exposed COD vials to light for three days before measuring.

Sample	Time (hours)	Chemical Oxygen Demand (mg/L)				Ethanol Concentration (mg/L)*				Nitrate Concentration (mg/L)				Nitrate Removal (%)		
		PTU	RAI	RAO	RBO	PTU	RAI	RAO	RBO	PTU	RAI	RAO	RBO	Reactor A	Reactor B	Total
S0	0	0	60	30	23	0.0	34.2	18.4	14.7	22.8	90.6	17.0	5.0	81.2	13.2	94.5
S1	1										83.0	14.8	4.0	82.2	13.0	95.2
S2	2		54.1	31	30		31.1	18.9	18.4		84.3	22.1	4.6	73.8	20.8	94.5
S3	3										90.3	21.9	5.2	75.8	18.5	94.2
S4	4		60	27	21		34.2	16.8	13.6		85.2	19.0	4.6	77.7	16.9	94.6
S5	5										88.8	18.0	3.9	81.9	13.7	95.6
S6	6		61	32	21		34.8	19.4	13.6		87.3	14.2	3.2	83.7	12.6	96.3
S7	7										89.1	13.4	2.5	85.0	12.2	97.2
S8	8		58	23	20		32.1	14.7	13.1		82.0	12.8	2.5	84.4	12.6	97.0
Average Values		0.0	58.2	28.6	23.0	0.0	33.3	17.6	14.7	22.8	86.7	16.8	3.9	80.6	14.8	95.5

Nitrate Removal, Flow = 1.0 gpm
Legend: —■— % Total Removal —◆— % Removal Reactor A —▲— % Removal Reactor B
y-axis: Nitrate Removal (%) — x-axis: Time (hours)

Nitrate Concentrations, Flow = 1.0 gpm
Legend: —◆— Influent Reactor A —■— Effluent Reactor A —▲— Effluent Reactor B
y-axis: Nitrate Conc. (mg/L) — x-axis: Time (hours)

Ethanol Concentrations, Flow = 1.0 gpm
Legend: —◆— Influent Reactor A —■— Effluent Reactor A —▲— Effluent Reactor B
y-axis: Ethanol Conc. (mg/L) — x-axis: Time (hours)

Nitrate Test Data

Test 3, Flow rate = 1.5 gpm = 5.7 Lpm

Sample	Time (hours)	Chemical Oxygen Demand (mg/L)				Ethanol Concentration (mg/L)				Nitrate Concentration (mg/L)				Nitrate Removal (%)		
		PTU	RAI	RAO	RBO	PTU	RAI	RAO	RBO	PTU	RAI	RAO	RBO	Reactor A	Reactor B	Total
S0	0	0	86	18	10	0.0	48.0	12.0	7.8	24.16	87.3	14.9	2.8	82.9	13.9	96.8
S1	1										85.2	28.4	3.9	66.7	28.8	95.4
S2	2		87	28	14		48.5	17.3	9.9		82.4	31.1	6.5	62.3	29.9	92.1
S3	3										81.7	30.9	7.3	62.2	28.9	91.1
S4	4		88	30	11		49.0	18.4	8.3		84.2	31.0	8.0	61.8	28.5	90.1
S5	5										82.8	31.7	8.0	62.4	28.1	90.5
S6	6		75	43	16		42.2	25.2	10.9		83.8	27.1	8.2	67.3	22.8	90.1
S7	7										85.6	30.9	8.0	63.1	27.3	90.5
S8	8		85	44	12		38.9	25.8	8.8			31.4	7.8	63.3	27.8	91.1
Average Values		0.0	80.2	32.6	12.6	0.0	44.9	19.7	9.1	24.2	83.7	28.6	5.7	65.7	26.2	92.0

Nitrate Removal, Flow = 1.5 gpm

— % Total Removal   — % Removal Reactor A   — % Removal Reactor B

Nitrate Concentrations, Flow = 1.5 gpm

— Influent Reactor A   — Effluent Reactor A   — Effluent Reactor B

Ethanol Concentrations, Flow = 1.5 gpm

— Influent Reactor A   — Effluent Reactor A   — Effluent Reactor B

Nitrate Test Data

Test 4, Flow rate = 2.0 gpm = 7.6 lpm

Sample	Time (hours)	Chemical Oxygen Demand (mg/L)				Ethanol Concentration (mg/L)				Nitrate Concentration (mg/L)				Nitrate Removal (%)		
		PTU	RAI	RAO	RBO	PTU	RAI	RAO	RBO	PTU	RAI	RAO	RBO	Reactor A	Reactor B	Total
S0	0	0	90	24	14	0.0	50.1	15.2	9.9	23.7	95.3	20.2	12.8	69.4	17.2	86.6
S1	1										85.3	40.9	13.1	52.0	32.6	84.8
S2	2		84	30	18		46.9	18.4	12.0		91.8	40.7	14.8	55.7	28.2	83.9
S3	3										88.5	42.0	15.8	52.5	29.6	82.1
S4	4		86	20	21		48.0	17.8	13.6		94.7	40.5	16.3	57.2	25.6	82.8
S5	5										92	41.1	18.0	55.3	27.3	82.6
S6	6		79	32	20		44.3	19.4	13.1		94.9	41.2	15.9	56.6	26.7	83.2
S7	7										89.8	38.4	14.5	57.2	26.6	83.9
S8	8		75	35	17		42.2	21.0	11.5		90.5	39.5	14.3	56.4	27.8	84.2
Average Values		0.0	82.8	30.0	18.0	0.0	46.3	18.4	12.0	23.7	91.4	39.3	14.8	56.9	26.8	83.8

Ethanol Concentrations, Flow = 2.0 gpm

Nitrate Concentrations, Flow = 2.0 gpm

Nitrate Removal, Flow = 2.0 gpm
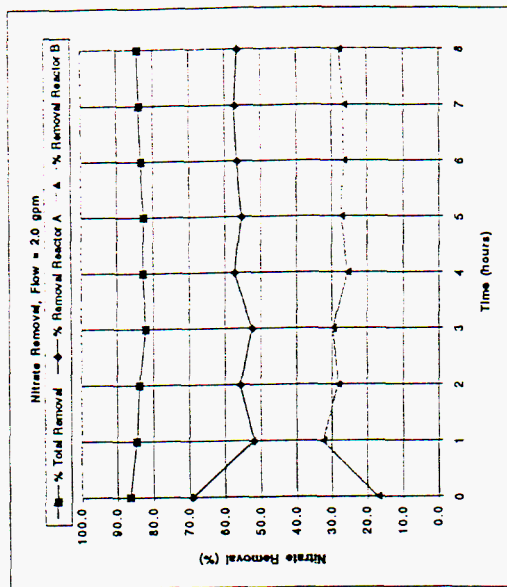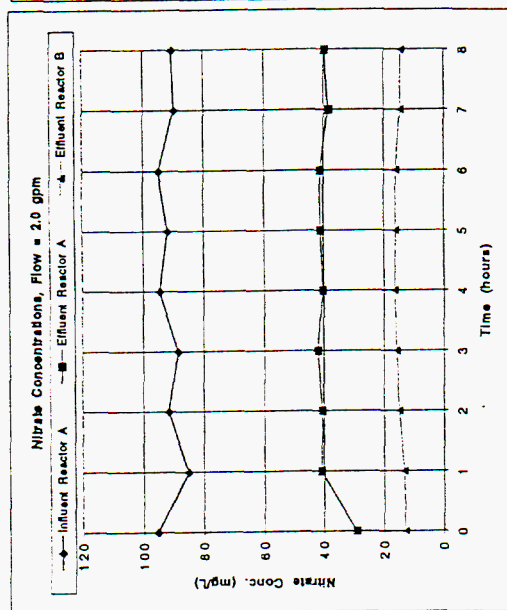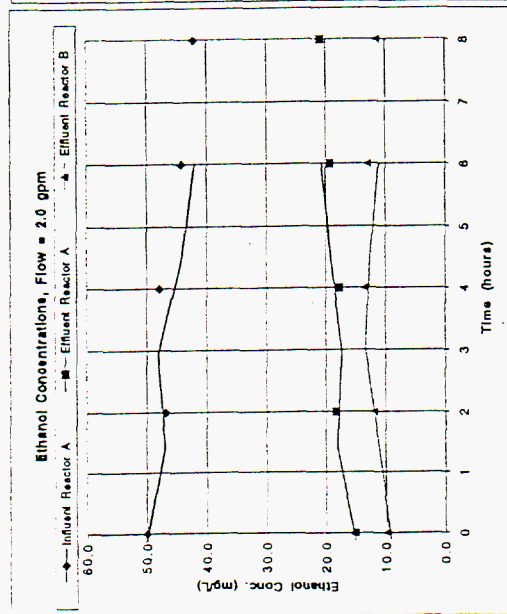
Perchlorate Study (8-10-99)

Nitrate Concentration (mg/L), 10^2 x Bromide Concentration (mg/L), Perchlorate Concentration (ug/L) vs. Time (hours)
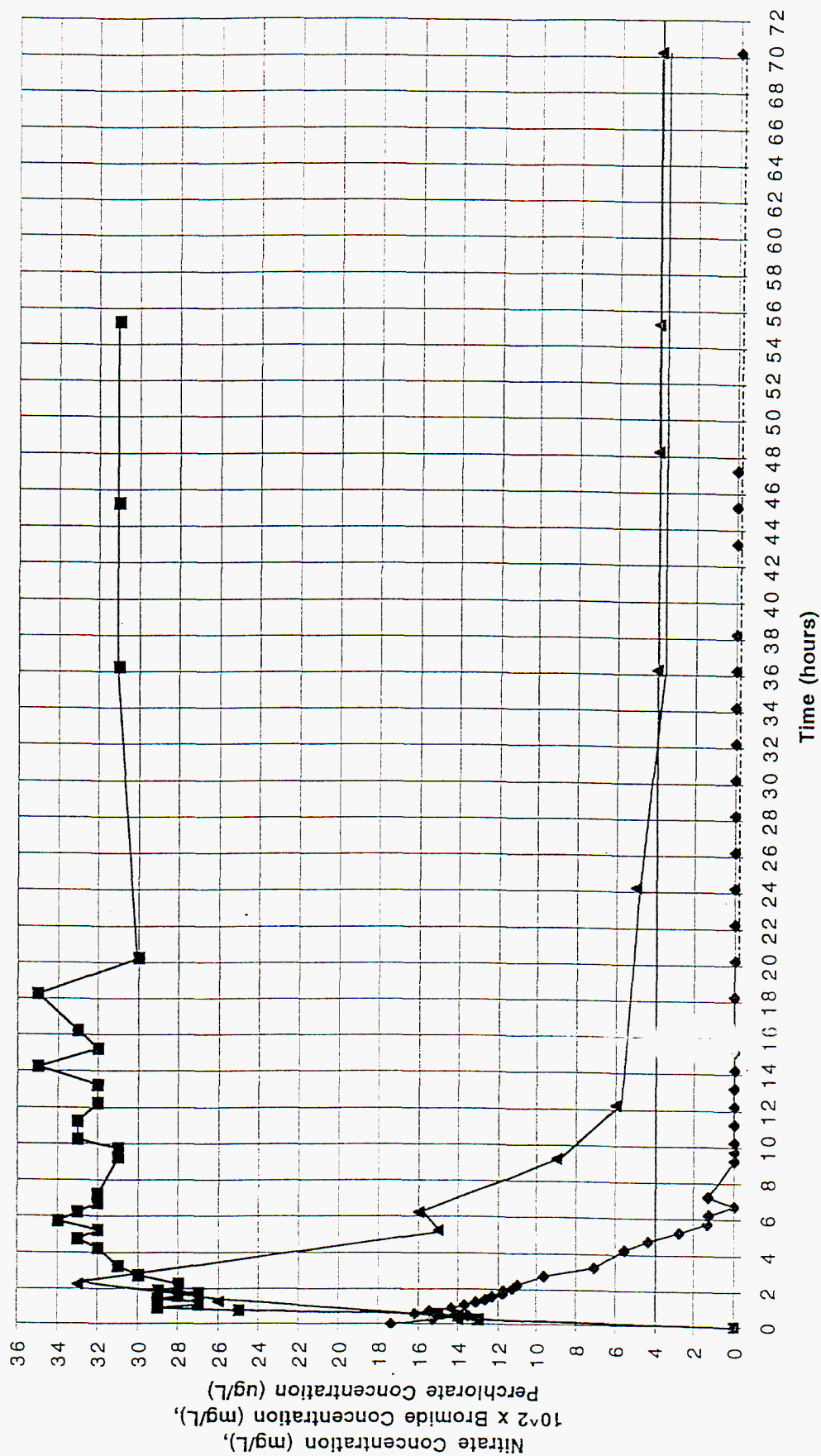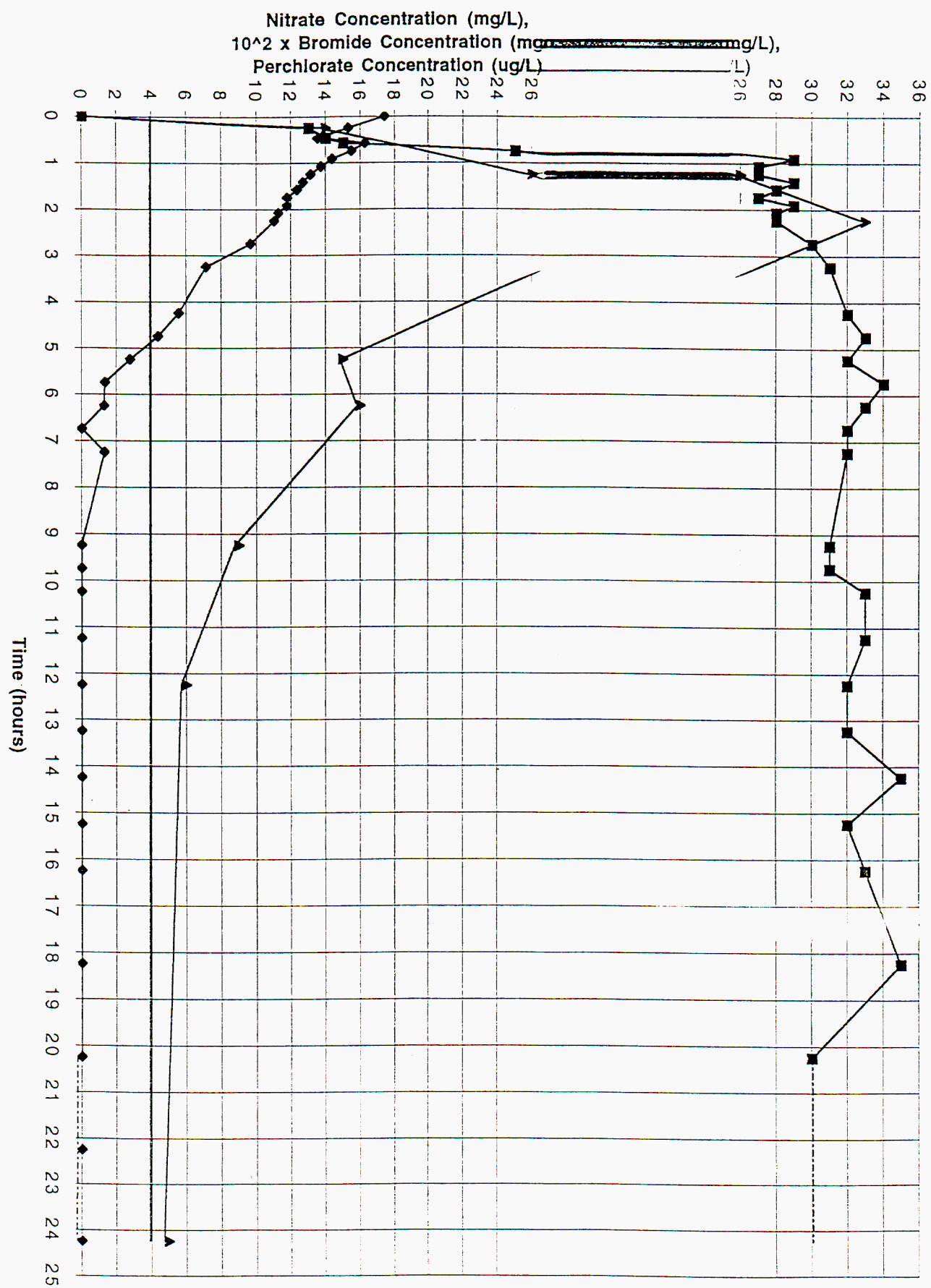
Appendix B
Kinetic study of denitrification

where:

A$_s$: is the total surface area inside the bioreactor

r' : is the rate of reaction with respect to surface area

Zero Order

The zero-order reaction rate is equal to the following:

$$r' = k_o'$$

Plug Eq. B-3 into Eq. B-2 and integrate.

$$\frac{A_s}{Q} = \frac{1}{k_o'}\left(C_{in} - C_{out}\right)$$

Where k_o' is the zero-order rate constant with units of *concentration/(ft/min)*.

First Order

The first-order reaction rate is equal to the following:

$$r' = k' * C$$

Plug Eq. B-5 into Eq. B-2 and integrate.

$$\frac{A_s}{Q} = \frac{1}{k'}\ln\frac{C_{in}}{C_{out}}$$

Where k' is the first-order rate constant with units of *ft/min*.

Second Order

The second-order reaction rate is equal to the following:

$$r' = (k_2')^2 * C$$

Plug Eq. B-7 into Eq. B-2 and integrate.

$$\frac{A_s}{Q} = \frac{1}{k_2'}\left(\frac{1}{C_{out}} - \frac{1}{C_{in}}\right)$$

Eq. B-8

Where k_2' is the second-order rate constant with units of *ft/(concentration*min)*.

Data

Table B-1: Summary of BTU-1 nitrate degradation results (from Appendix A)

Flow Rate gpm	Cin mg/l	Cout,A mg/l	Cout,B mg/l
0.5	99.9	22.8	7.1
1.0	86.7	16.8	3.9
1.5	83.7	28.6	6.7
2.0	91.5	39.4	14.8

Note: Total surface area of each reactor is 770 ft^2 (see Appendix E for estimation)

Schematic of BTU-1
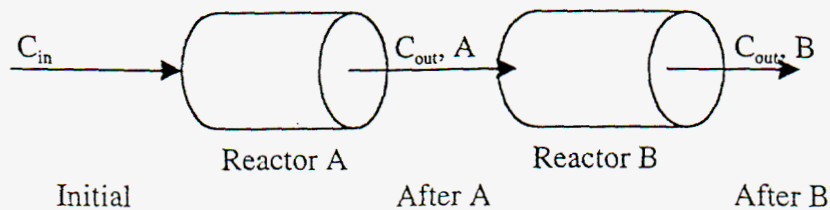
Figure B-1: BTU-1 consists of two horizontal tanks (500 gallons each) with corrugated plates inside to provide a surface for microorganisms to grow and form biofilms. Nitrate concentrations were measured at various stages: initial (influent), reactor A effluent, and reactor B effluent.

Data Analysis

Table B-2: Data used for reaction order plots

Flow rate gpm	stage	Nitrate conc. mg/L	Surface Area ft²	A_s/Q ft²/gpm	zero order $C_{in} - C_{out}$ mg/L	1st order ln (C_{in}/C_{out})	2nd order $1/C_{out} - 1/C_{in}$ L/mg
0.5	initial	99.9	0	0	0.0	0.000	0.000
0.5	after A	22.8	771	1542	77.1	1.477	0.034
0.5	after B	7.1	1542	3084	92.8	2.644	0.131
1.0	initial	86.7	0	0	0.0	0.000	0.000
1.0	after A	16.8	771	771	69.9	1.641	0.048
1.0	after B	3.9	1542	1542	82.8	3.101	0.245
1.5	initial	83.7	0	0	0.0	0.000	0.000
1.5	after A	28.6	771	514	55.1	1.074	0.023
1.5	after B	6.7	1542	1028	77.0	2.525	0.137
2.0	initial	91.5	0	0	0.0	0.000	0.000
2.0	after A	39.4	771	385.5	52.1	0.753	0.014
2.0	after B	14.8	1542	771	76.7	1.822	0.057
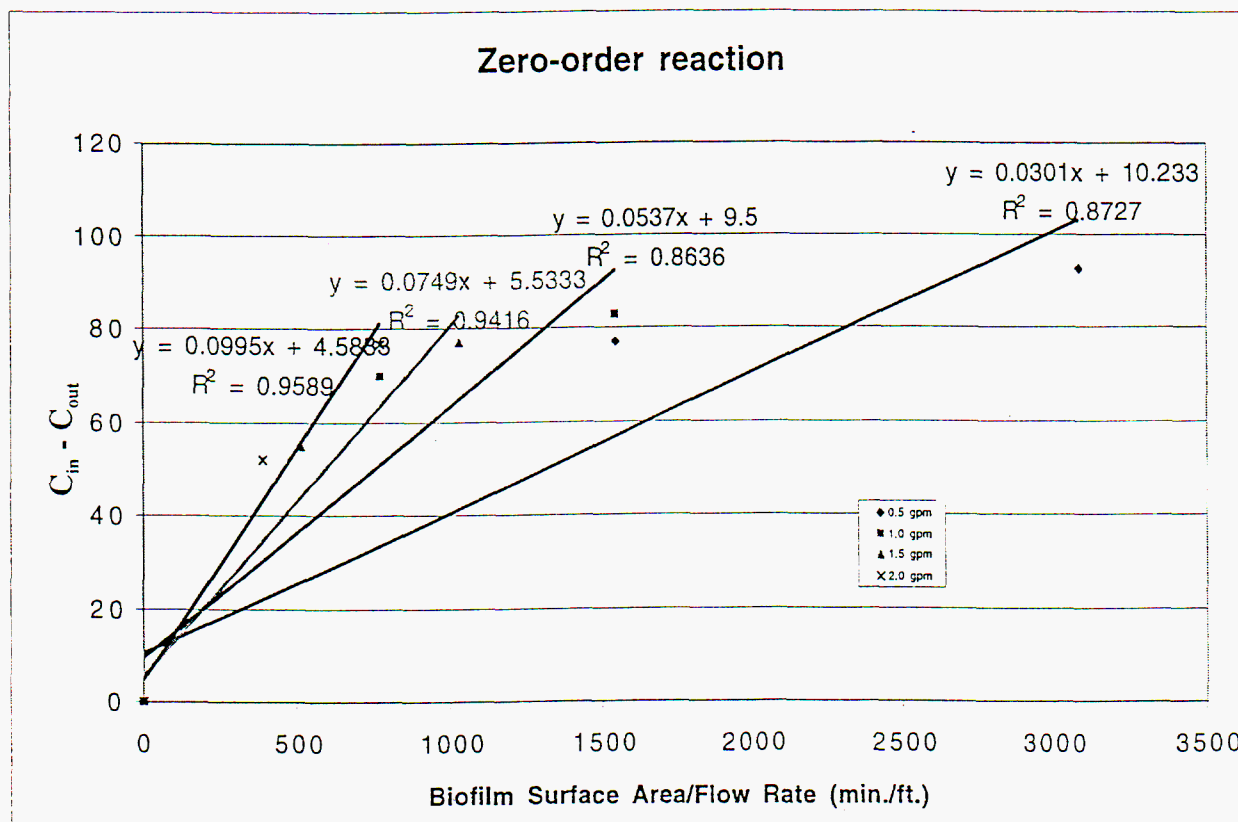
Figure B-2: Linear regression of zero-order reaction.

First-order reaction

$y = 2.01\text{E-}03x + 3.02\text{E-}02$
$R^2 = 9.99\text{E-}01$

$y = 2.46\text{E-}03x - 6.28\text{E-}02$
$R^2 = 9.93\text{E-}01$

$y = 2.36\text{E-}03x - 5.27\text{E-}02$
$R^2 = 9.90\text{E-}01$

$y = 8.57\text{E-}04x + 5.17\text{E-}02$
$R^2 = 9.95\text{E-}01$

$\ln(C_{in}/C_{out})$

Biofilm Surface Area/Flow Rate (min./ft.)

- ◆ 0.5 gpm
- ■ 1.0 gpm
- ▲ 1.5 gpm
- ✕ 2.0 gpm
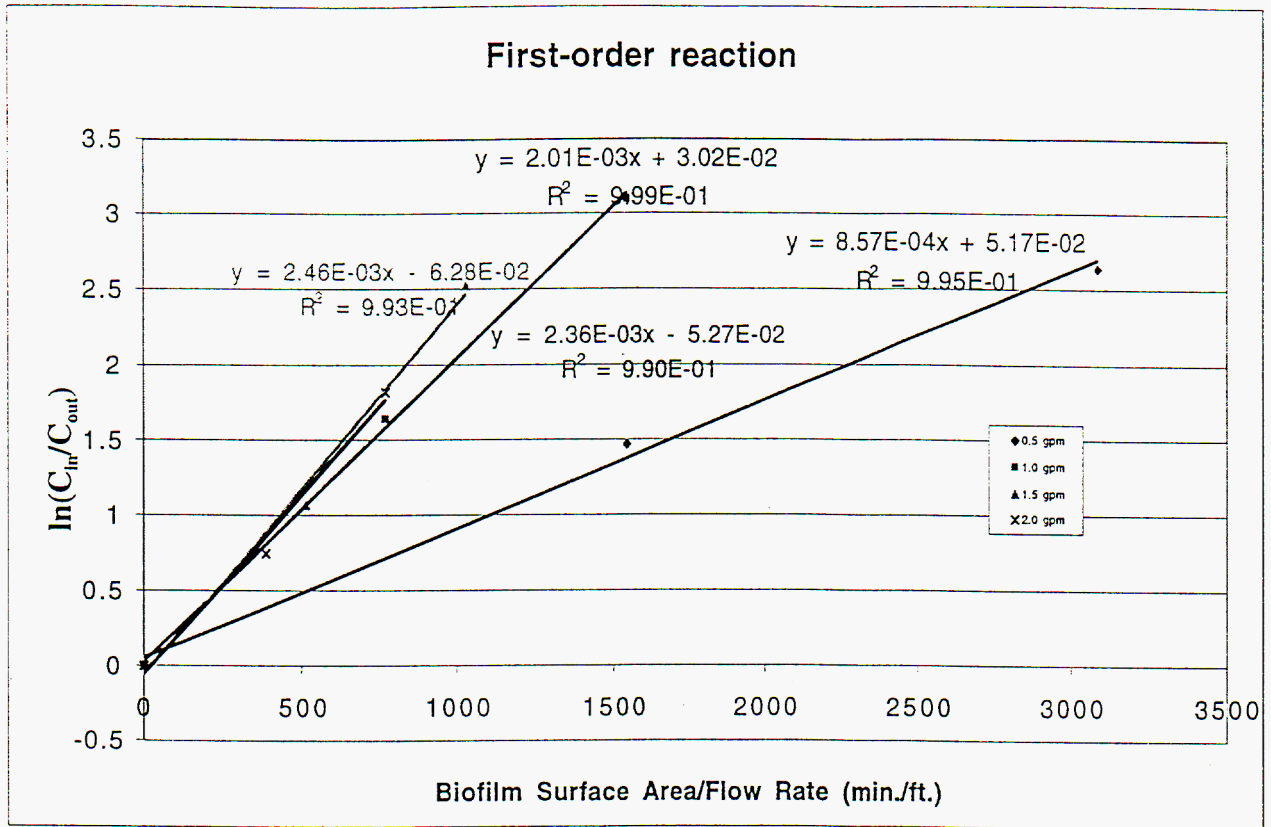
Figure B-3: Linear regression of first-order reaction.

Second-order reaction

$y = 0.0002x - 0.0248$
$R^2 = 0.8902$

$y = 0.0001x - 0.0152$
$R^2 = 0.8718$

$y = 4E\text{-}05x - 0.0105$
$R^2 = 0.9284$

$y = 7E\text{-}05x - 0.0048$
$R^2 = 0.9206$

- ◆ 0.5 gpm
- ■ 1.0 gpm
- ▲ 1.5 gpm
- ✕ 2.0 gpm

Y-axis: $1/C_{out} - 1/C_{in}$

X-axis: Biofilm Surface Area/Flow Rate (min./ft.)
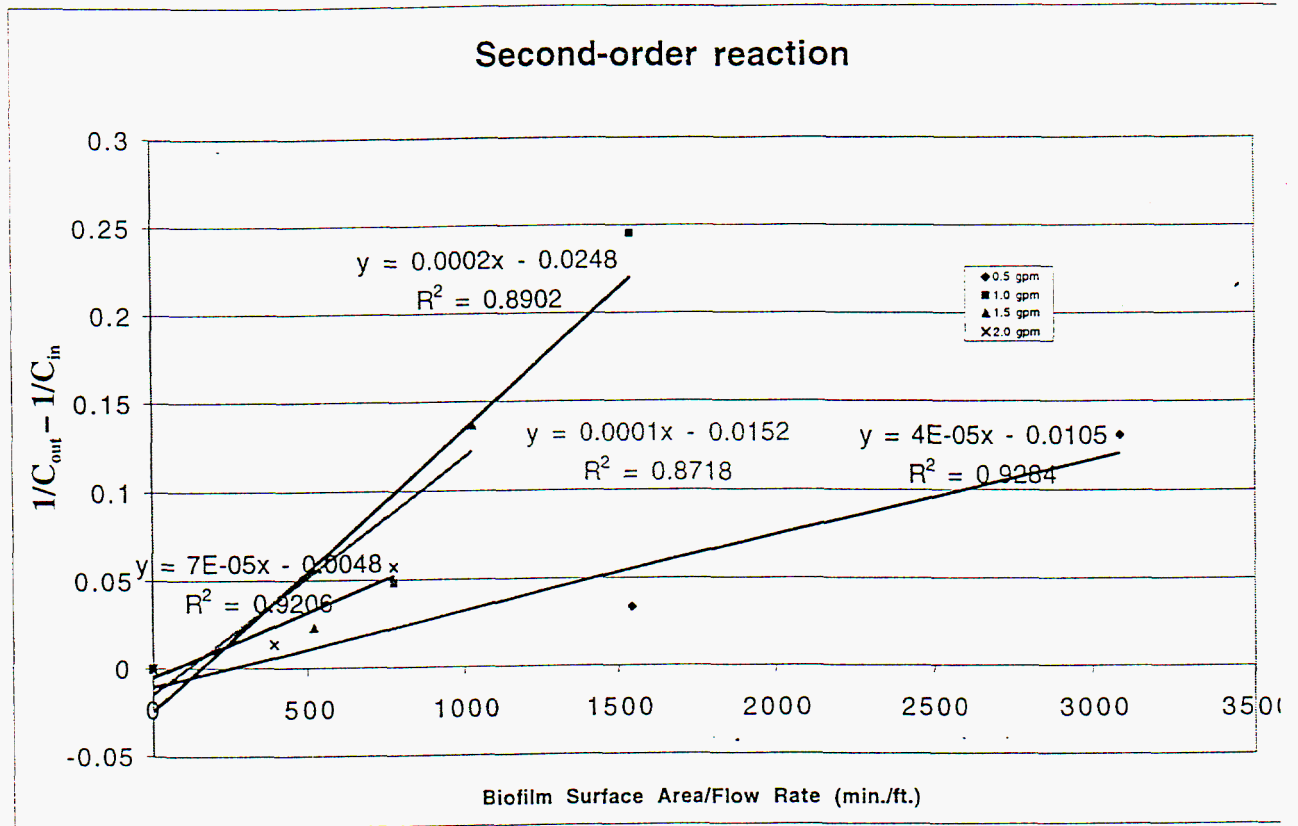
Figure B-4: Linear regression of second-order reaction.

Results

Table B-3: Summary of linear regression results

Flow rate	First order		Second order		Zero order	
gpm	Slope	R²-value	Slope	R²-value	Slope	R²-value
0.5	0.00236	0.990	0.0000739	0.921	0.0995	0.9589
1	0.00246	0.993	0.0001590	0.89	0.0537	0.9416
1.5	0.00201	0.999	0.0013300	0.872	0.0749	0.8636
2	0.00086	0.995	0.0000425	0.928	0.0301	0.8727

From **Figures B-2** through **B-4** and R^2-values in the above table, we conclude that the denitrification reaction can best be modeled as a first-order process.

Appendix C
Perchlorate study

In the first two hours of the experiment, the measured concentration of perchlorate increased from 4 to 33 μg/L (ppb) as seen in **Figure C-1**. The increase in perchlorate concentration was probably caused by incomplete mixing. When the time approached the 12-hour mark, the perchlorate concentration approached the detection limit of 4 ppb. For modeling purposes, only data reflecting initial conditions after complete mixing were considered (**Figure C-2**).

Exponential Expressions

$y = 29.58e^{-0.1654x}$
$R^2 = 0.9686$

$y = 12.066e^{-0.5017x}$
$R^2 = 0.9244$

Time (hours)

Nitrate Concentration (mg/L)
Perchlorate Concentration (μg/L)
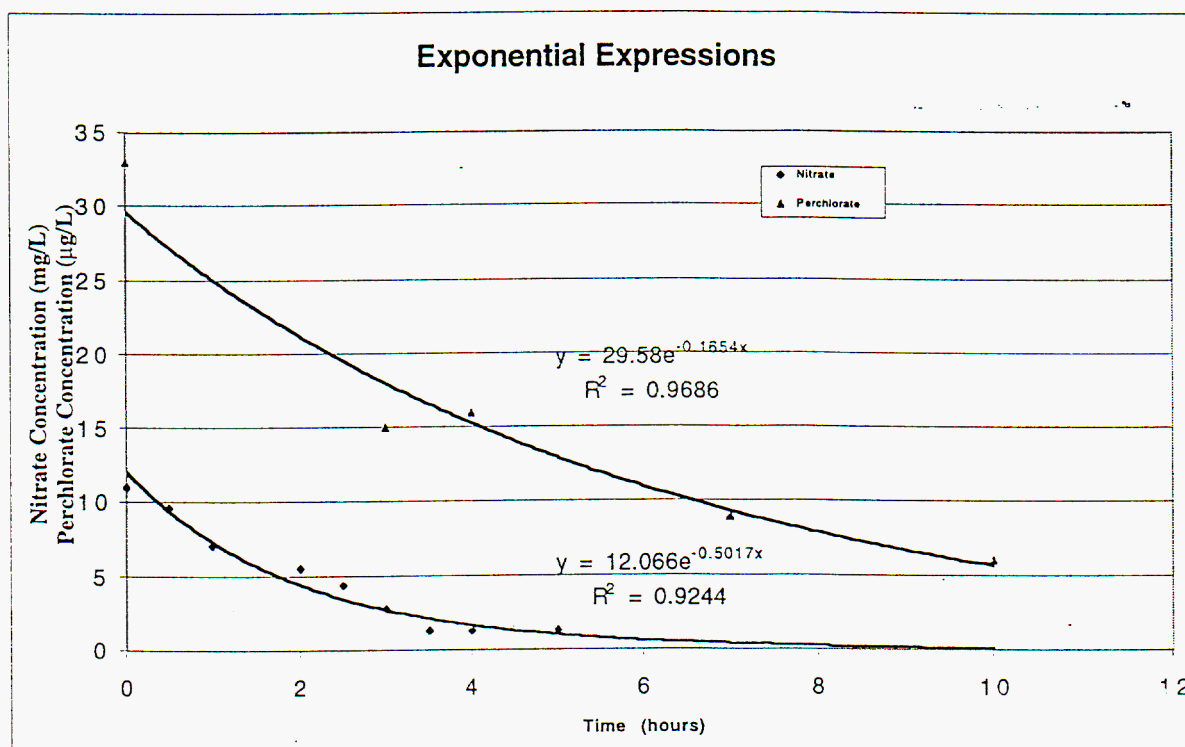
Figure C-2: Modified data for the BTU-1 batch study.

Data Analysis

In **Figure C-2** we see that both nitrate and perchlorate concentrations decrease exponentially. To find the relationship between the two degradation rates, we use exponential trends to express the kinetics:

$$C = C_{in} * e^{-kt}$$

<div align="right">Eq. C-1</div>

where:

C: is the concentration at a given time

C_{in}: is the initial concentration

t: is the time (hr)

k: is the reaction rate constant (hr^{-1})

The rate constant, k, was found to be 0.50 hr^{-1} and 0.17 hr^{-1} for nitrate and perchlorate, respectively. The ratio of the k value for nitrate to perchlorate is 3; thus perchlorate degradation was slower and proceeded at a rate 1/3 of that for nitrate degradation.

Appendix D
Denitrification temperature dependence

Purpose: This section is concerned with the influence of temperature on the rate of denitrification in the new bioreactor. Predictions for the bioreactor performance must be adjusted for well water temperature (18 °C).

Methodology: The data used in this study were collected by A. L. Parker at the Department of Chemical Engineering, University of Wisconsin (Parker, 1976). This experiment was performed at varying operating temperatures in packed beds similar to those found in the bioreactor. A first-order rate expression applied.

Data

Table D-1. Modeling of low NO_3—N data with a first-order kinetic model

Temperature C	k (hr-1)	Res. Stnd. Error (mg/L)	Obs. Stnd. Dev. (mg/L)
20	0.463	7.73	7.95
13	0.311	3.34	3.11
5	0.236	0.80	0.73

Data Analysis

Over the range of temperatures studied (5 to 20°C), the Arrhenius equation should be suitable for modeling the temperature dependence:

$$k = Ae^{-u/RT}$$

Eq. D-1

Taking the natural log of Eq. D-1, we have the following:

$$\ln k = \ln A - \frac{u}{RT}$$

Eq. D-2

Plotting ln(k) vs. 1/T we obtain a linear correlation as seen in **Figure D-1**.
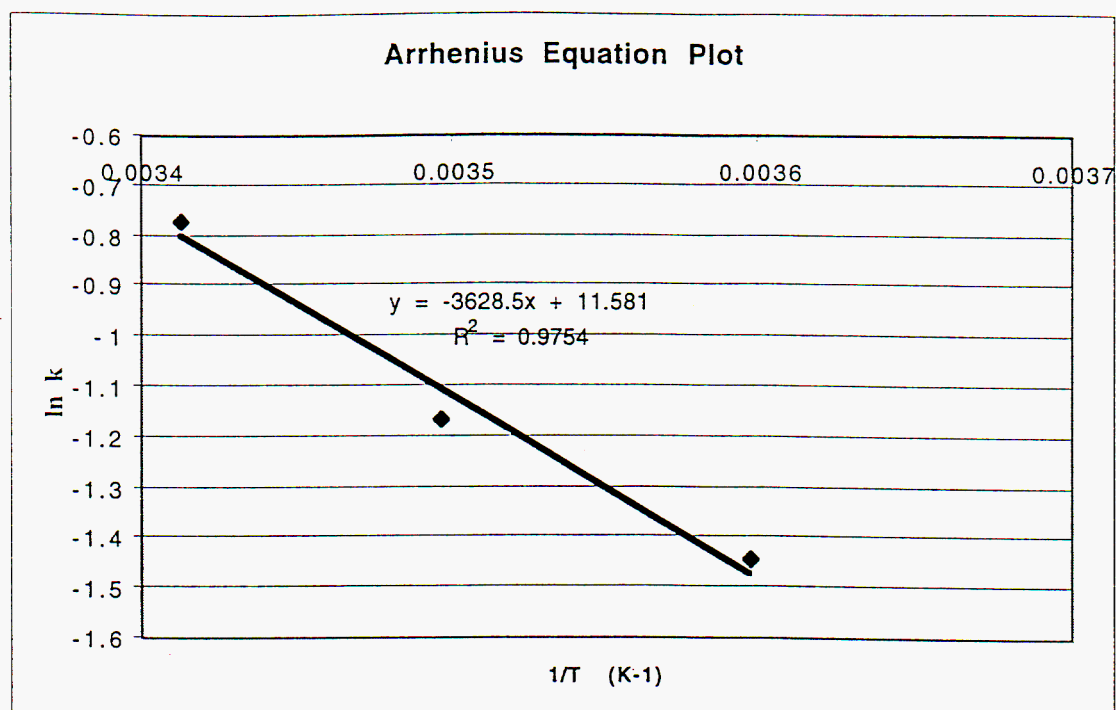
Figure D-1: This plot expresses the linear relationship of ln(k) and 1/T.

The plot in **Figure D-1** indicates that the temperature dependence of nitrate degradation can be expressed with the Arrhenius Equation. A plot of the first-order rate constant versus the operating temperature is informative (see **Figure D-2**).
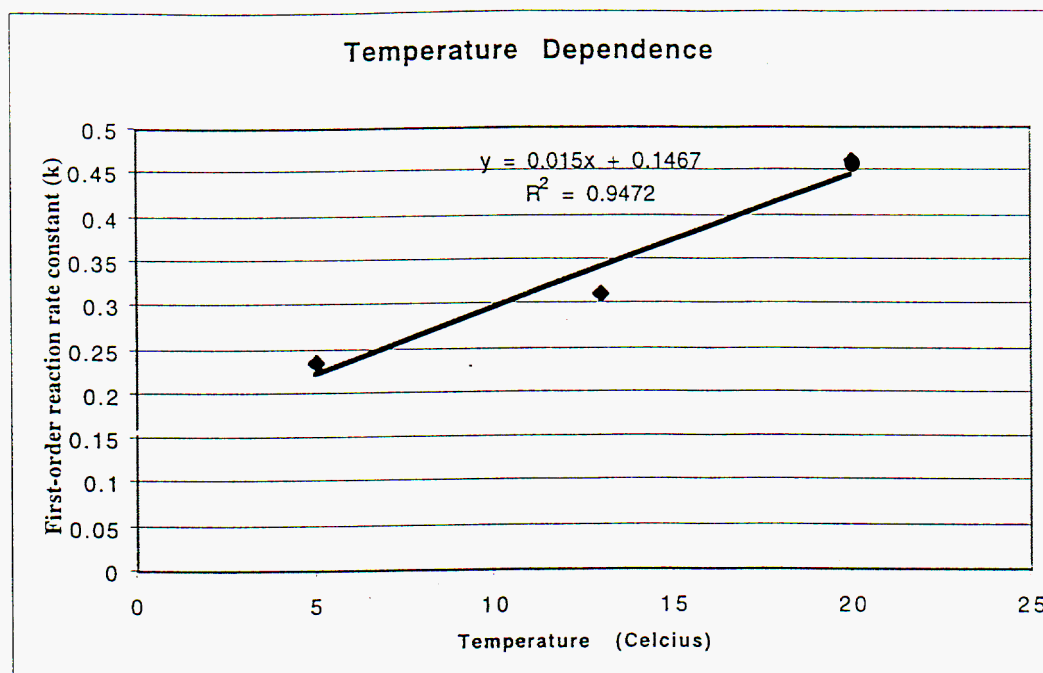
Figure D-2: The rate constant k increases linearly with temperature.

Conclusions

By examining **Figure D-2**, we observe that a decrease in the water temperature by 10 °C reduces the first-order rate constant (k) by a factor of 1.5. To keep our predictions conservative, we will assume that a 10 °C decrease will result in a reaction rate constant decrease of a factor of 2.

Appendix E
Sample bio-tower calculations

Estimation of Packing Surface Area 9-3-99

For Btu-1

Surface Area to volume ratio - for current unit

spacing -

Suspended Biomass?
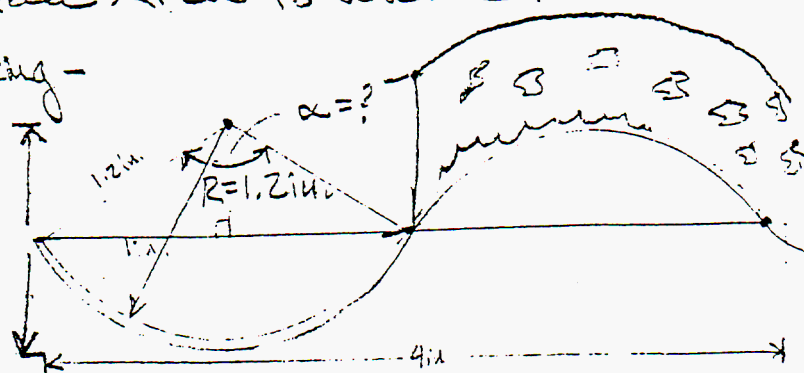
plates @ 2" o.c.
thickness = 1/16"
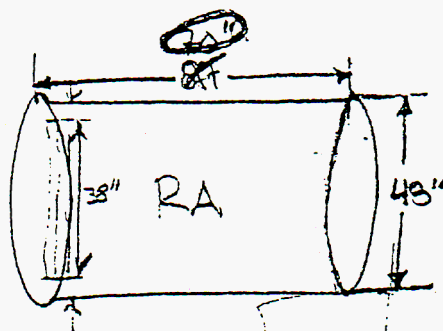
$$\frac{1}{2}\alpha = \sin^{-1}\frac{o}{h} = \sin^{-1}\left(\frac{1}{1.2}\right) = 56.4°$$

$$\alpha = 56.4 \times 2 = \underline{112.8°}$$

part of circle, $\dfrac{112.8°}{360°} = 0.31357$

circumference:

$$C = 2\pi R = 2 \cdot 3.14 \cdot 1.2 in.$$
$$C = 7.536 in.$$

$$(\cup)\ L = 0.314 \cdot 7.536 = \underline{2.363 in.}$$

length of $(\sim)$ = $2 \cdot 2.363 in. = \underline{4.73 in.}$

70"/2 = 35 plates

38"/4" = 9.5 ($\sim$) curves

$$L_{total} = 9.5 \cdot 4.73 in. = \underline{44.935 in.}$$

$$A_{plate} = \frac{\pi (44.935 in)^2}{4} = 1,585 in^2.$$

$$A_{total} = 1,585 in^2 \cdot 35 = 55,476 in^2 \quad \frac{1 ft^2}{144 in^2}$$

$$\boxed{A = 385 ft^2} - WS\ reactor\ packing$$

Surface Area $\sim 385 ft^2 \cdot 2 \sim \boxed{770 ft^2 = A_s}$

Appx. Volume $(V) = 1,585 in^2 \cdot 2 in. = 3170 in^3 \quad \dfrac{ft^3}{(12 in)^3} \cdot 35$

$$\boxed{V \sim 64 ft^3}$$

$$A_s/V\ ratio = \frac{770 ft^2}{69.6 ft^3} \cdot \frac{ft^3}{7.48 gal} \cdot \frac{gal}{3.785L} \cdot \frac{m^2}{(3.28 ft)^2} \cdot \frac{(100 cm)^2}{m^2}$$

$$\boxed{A_s/V = 394\ cm^2/L}$$

-Sean Spargo

Column Calculations

BioReactor	old BTU-1	new#1 BTU-2	new#2 BTU-2.1	units	EXPLANATIONS:
Number of Towers	2	2	3		
Diameter:	48	30	30	inches	
Tower Length/Height:	117	16	24	feet	BTU-2(&2.1)'s height is the addition of two towers run in series
Reactor Packing Fraction:	?	0.75	0.75		% of Reactor that is filled with packing
Average Temp.:	23	18	18	°C	
Reactor Volume:	146.6	78.5	117.8	ft³	Calculated from the dimensions
	1097	588	881	gallons	
Bed (Packing) Volume:	961	441	661	gallons	Volume occupied with packing
Wall Thickness:	?	0.25	0.25	inches	
Steel Density:	487	487	487	lb/ft³	
Steel Weight:	?	640	960	lb	
Packing S.A./Vol.:	12	85	85	ft²/ft³	BTU-1 had plates, BTU-2(&2.1) has 1" Jaeger Tri-Pack
Packing S.A.:	1542	5007	7510	ft²	S.A/Vol. Ratio Multiplied by Bed Volume
Packing Weight:	?	280	420	lb	Density of Packing Multiplied by Bed Volume
Packing Void:	?	0.9	0.9		Packing Space that is unoccupied by plastic
Water Volume:	?	543	815	gallons	
Water Weight:	?	4535	6803	lb	
BioReactor Weight:	?	5455	8183	lb	Neglects additional tower equip. (Piping, Top & Bottom etc.)
Treatment Flow:	2	5	5	gpm	
Packing S.A.:	1542	5007	7510	ft²	
Nitrate In:	91.4	90	90	mg/L	BTU-1 is an average over a day's time, BTU-2(&2.1) value is standard
Retention Time:	9.1	2.0	2.9	hr	Calculated from Reactor Volume/Treatment Flow
Hydraulic Load:	0.021	0.136	0.136	ft/min	
k $(Q/V_T)*\ln(C_o/C_i)$	0.0038	0.0038	0.0038	min⁻¹	BTU-1 k value calculated from Nitrate Effluent data, BTU-2(&2.1) k set equal
Nitrate Out:	14.8	64.	54.5	mg/L	BTU-2,(&2.1) Nitrate Effluent Solved for from 'k' value
k' $(Q/A_T)*\ln(C_o/C_i)$	0.00032	0.00032	0.00032	ft/min	BTU-1 k' value calculated from Nitrate Effluent data, BTU-2(&2.1) k' set equal
Nitrate Out:	14.8	8.	2.5	mg/L	BTU-2(&2.1) Nitrate Effluent Solved for from k' values
k' (Temp. Correction):	0.00032	0.00024	0.00024	ft/min	Correction of k' for temp. (reduce by factor of 2 for 10°C decrease)
Nitrate Out (Temp. Correction):	14.8	14.9	6.0	mg/L	

k' is a normalized value using packing S.A. at a correspondence to catalyst weight, instead of using volume

underline: values from Shawn's calculations

Appendix F
Flammability of ethanol in water

Flammability of Ethanol

Table F-1: Thermodynamic properties of saturated ethanol

Temperature	Temperature	Temperature	Vapor pressure, p^{vap}	Vapor pressure, p^{vap}
K	°C	°F	bar	atm
270	-3	27	0.0128	0.0130
280	7	45	0.0250	0.0253
290	17	63	0.0480	0.0486
300	27	81	0.0880	0.0892
310	37	99	0.1510	0.1530
320	47	117	0.4060	0.4114

Perry, Robert H., Green, Don W. Perry's Chemical Engineers' Handbook. 7th edition. McGraw Hill, 1997. p. 2-235.

Table F-2: Ethanol's flammable limits

	% by volume	partial pressure (atm)
lel (lower explosive limit)	3.3	0.033
uel (upper explosive limit)	19	0.19

$$p = y_i P = x_i p^{vap}$$ Eq. F-1

p: partial pressure of ethanol in air
y_i: mole fraction of ethanol in air
x_i: mole fraction of ethanol in water
P: total pressure (1atm)
p^{vap}: vapor pressure of ethanol

Table F-3: Ethanol flammability data

Temperature °F	Vapor pressure, p^{vap} bar	molar fraction of ethanol in water	% volume of ethanol in water
26.6	0.0128	1.00	100%
44.6	0.025	1.00	100%
62.6	0.0480	0.69	81%
80.6	0.0880	0.38	53%
98.6	0.1510	0.22	35%
116.6	0.4060	0.08	14%
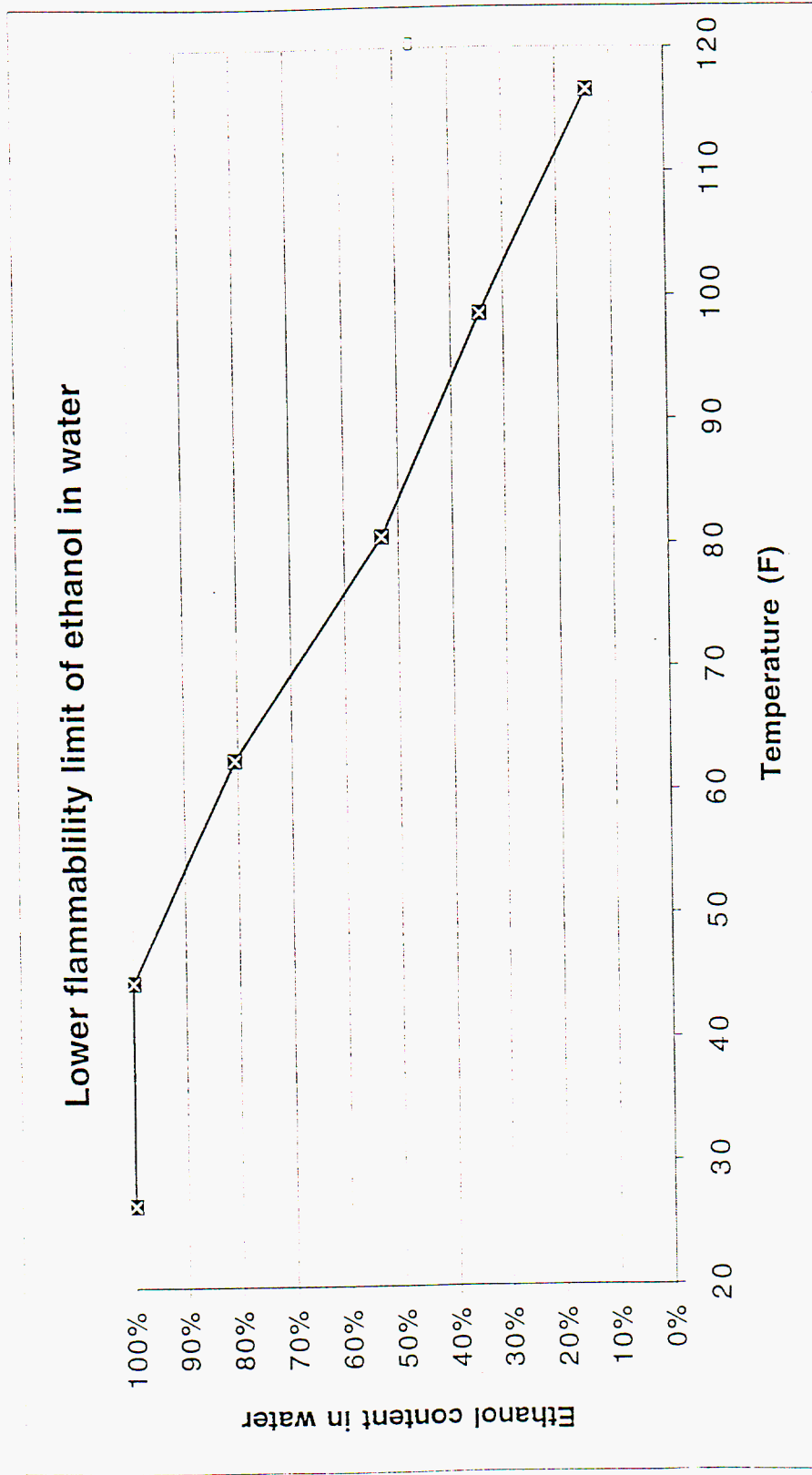
Figure F-1: Flammability limit of aqueous ethanol solutions as a function of temperature. Safe operation of the reactor will require feed solutions with an ethanol content of less than the lower flammability limit.

The Prediction of Biochemical Acid Dissociation Constants Using First Principles Quantum Chemical Simulations

Ngoc L. Tran and Michael E. Colvin*

Computational Biology Group, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94550

Abstract: Proton transfer is a vital part of many chemical processes and is determined by the acid dissociation constants (pK_a) of the chemicals involved. The goal of this study is to evaluate existing quantum chemical methods to accurately predict pK_a and to determine the trade-off between accuracy and computational cost. We used density functional theory (DFT) with the B3LYP functional and two basis sets, 6-31G** and 6-31++G(3df,3pd). To include the effects of aqueous solvation, we used DFT combined with a conductor-like screening solvation model as well as the Langevin dipole model. Using largest basis set, aqueous-phase optimized structures, we find a strong linear relationship between the predicted and calculated pK_a's (R^2=0.94) and even stronger fits within the individual classes of compounds. Despite these strong linear correlations, we find a systematic error leading to a much larger range in the predicted pK_a values. We studied the effect of excluding the enthalpy and entropy terms, using the gas-phase optimized structures, and using a smaller basis set, and found in all cases no significant decrease in the accuracy of the predicted pK_a values. We also calculated the pK_a values from the gas-phase energies, excluding any solvation effects, and also found a strong linear relation with the experimental pK_a values, although with much lower systematic errors in the range of predicted pKa values. Finally, we found that the Langevin dipole method yielded pKa values with smaller absolute errors than the PCM methods, but yielded poorer linear fits to the experimental values.

Introduction

In water, many chemicals can readily gain or lose protons (H$^+$ ions). This transfer of protons between molecules is a vital part of many chemical processes and plays a central role in many biochemical reactions. For example, most biosynthetic and biodegradation reactions are dehydration or hydrolysis involving proton transfer. On a larger scale, proton transfer is coupled to many biological processes, such as the transport of Na$^+$ and Ca^{2+} in the cell membrane as well as the synthesis of adenosine triphosphate (ATP).

The propensity to gain or lose protons is defined by the chemical's acid dissociation constant (pK$_a$). The pK$_a$ is an innate property of a molecule that describes its protonation state under different pH conditions. The pH is a property of an aqueous solution defined as -log[H$^+$] where [H$^+$] is the proton concentration. The precise relationship between pH, pK$_a$, and the fraction of deprotonated acid is given by the Henderson-Hasselbalch equation [1]:

$$pH = pK_a + \frac{[A^-]}{[AH]} \tag{1}$$

The ability to calculate pK$_a$ values is very useful because experimental measurement can be expensive or difficult to obtain. The experimental measurement of pK$_a$ by simple titration does not provide information about the site of protonation. For example, phosphoramide mustard, an active metabolite of the anti-cancer drug cyclophosphamide, has three potentially protonable sites, so that determination of the actual protonation occurring at a given pH required NMR analysis of isotopically labeled phosphoramide mustard [2]. More generally, the ability to accurately calculate the energetics of proton loss or gain would increase the utility of quantum chemical simulations for studying aqueous-phase phenomena.

The accurate first principles calculation of pK$_a$ values is very difficult. The loss or gain of a proton inherently involves ionic species, so that solvation effects must included to cancel the very high gas-phase protonation and deprotonation energies. Moreover these large solvation energies must be calculated very accurately since the extent of typical pK$_a$ values covers an energy range of

only about 10 kcal/mole. Finally, the use of a finite basis set introduces errors in the relative gas-phase energies of charged and uncharged molecules (e.g. an acid and its conjugate base).

Ideally, solvation effects would be determined from molecular dynamics (MD) calculations wherein the time evolution of the solute and solvent is simulated. However, the large number of solvent particles required and the long simulation times prohibit using accurate *ab initio* force fields for these simulations. Instead, empirical forces are used which are not in general sufficiently accurate to predict quantative solvation energies. Another approach, the "supermolecule" methods, involves including a collection of fixed solvent molecules in the quantum chemical calculation. Although this method can be effectively used to probe the effect of specific solute-solvent interactions, it is inadequate for calculating total solvation energies because it fails to include dynamical properties of the solvent. .

An alternative to MD and supermolecule methods is to use a simplified model of the solvent, which implicitly includes the dynamical features of the solvent. There are a large number of implicit solvent models with different strengths and computational costs that have been described in an extensive review.[3] The conceptually simplest are those that model the interaction of a continuous dielectric medium with an analytical expression of the electrostatic moments of the solute. These expansions can be truncated at the monopole term (Born model[4, 5]), the dipole term (Onsager model[6, 7]), or higher expansions[8]. Additionally, these expansion methods can be generalized and combined with empirical corrections to produce very accurate solvation energies, such as the SM*x* series of methods developed by Cramer and Truhlar that average errors of less than 1 kcal/mole for neutral solutes and 3 kcal/mole for ions.[9] The drawback to these approaches, like all empirical methods, is that large errors may arise for compounds not included in the original parameterization.[10]

There are other types of implicit solvent models than numerically solve for the solvent effects in terms of a large set of point charges on a three-dimensional grid around the solute by solving the Poisson-Boltzman equations[11] or by solving the electrostatic continuity equations at the two dimensional boundaries between regions of different dielectric values (typically the solute-solvent

interface). The latter method, usually described as the polarizable continuum method (PCM), has been implemented in a number of forms with, for example, different choices for the definition of the solvent-accessible surface area that defines the solute-solvent dielectric boundary (see recent review[12]) Although these models exclude many important properties of water such as explicit hydrogen bonding and size effects, such methods have been shown to be surprisingly accurate at calculating solvation energies, solvent-phase geometries, and other properties that do not depend strongly on explicit molecular solvent effects. (Note however that there have been studies published describing the inherent shortcomings on PCM models[13, 14].)

A method related to these polarizable continuum model, but including a more realistic representation of the polar solvent, is the Langevin dipole method of Warshel, which models the solvent as a large set of polarizable dipoles on a fixed three-dimensional grid. This approach has recently been parameterized for use with *ab initio* derived solute charges and shown to yield solvation energies for neutral and ionic molecules comparable or better than PCM methods.[15]

The purpose of this paper is to evaluate the accuracy of several quantum chemical methods to calculate pK_a and to determine the trade-off between the methods' accuracies and computational costs. In this study, we used 10 compounds to evaluate the first principles prediction of pK_a values, four carboxylic acids, four imidazoles, phosphoric acid, guanine (see Figure 1.).

Methods

The pK_a can be calculated from the energy of the following reaction:

$$AH \xleftarrow{\Delta G} A^- + H^+ \qquad (2)$$

Using the relation,

$$pK_a = \frac{\Delta G}{2.303RT} = \frac{\Delta G}{1.36}$$

for T=298K. ΔG is the free energy of reaction obtained by taking the difference of the molecular free energies of the products from the reactants. Because of the large errors associated with comparing

the energies of neutral and ionic species and the uncertainties in the solvation energy of the isolated proton[16], it is generally much more accurate to calculate relative pK$_a$ values, using a reference compound (BH) with known pK$_a$ that is chemically similar to the compound under investigation (AH).

$$AH + B^- \xleftarrow{\Delta\Delta G} A^- + BH$$

$$pK_a(AH) = pK_a(BH) + \frac{\Delta\Delta G}{1.36} \tag{3}$$

These free energies consist of two parts, a gas-phase molecular free energy and a solvation free energy. *Ab initio* quantum chemical methods provide a large hierarchy of methods for calculating gas-phase electronic energies (ΔE) and molecular structures [17]. This hierarchy involves both a "level of theory" that determines the treatment of electron-electron interactions, and a basis set that determines the spatial flexibility in the construction of the molecular orbitals. Combining the electronic energy (E) with the zero-point energy (ZPE), thermal corrections to the enthalpy and entropy yields the gas-phase molecular free energy (G). These additional energy terms can be calculated using standard idealized statistical formulae from the harmonic vibrational frequencies [18].

We used ten compounds to evaluate different methods to calculate pK$_a$: four carboxylic acids, four imidazoles, four protonation states of phosphoric acid, and five protonation states of guanine (see Figure 1.). The gas phase structures and energies of all compounds studied were calculated using density functional theory (DFT) with the Becke 3-parameter hybrid exc   · f··n··   ··n! [19] and Lee-Yang-Parr gradient corrected electron correlation functional [20] (B3LYP), a method widely shown to accurately predict gas phase reaction energies, with a 6-31G** basis set and the much larger 6-31++G(3df,3pd) basis. Analytic B3LYP/6-31G* frequencies were calculated at the B3LYP/6-31G* optimized structures to determine the zero-point energies and thermal corrections to the enthalpy and the entropy. To determine the effect of aqueous solvation on the energy and structure, the compounds were also optimized using DFT combined with a derivative of PCM. The derivative, a conductor-like screening solvation model (COSMO) [21] models the surrounding solvent by means of polarization charges distributed on the solvent exposed surface of the molecule.

We also compared COSMO with another model that treats the solvent molecules like point dipoles (Langevin dipole) [15]. Following protocol, LD iterated and non-iterated were calculated from a structure derived from electronic energies calculated with the gas-phase HF/6-31G* at the gas-phase HF/6-31G* structure.

In this study, we evaluated the accuracy of several combinations of these methods for calculating pK$_a$'s, specifically comparing the alternative methods summarized in Table 1.

Results and Discussion

We will begin by analyzing the accuracy of the most computationally costly method for calculating molecular energies evaluated in this study. This involved using free energies (G) derived from large basis set CPCM-B3LYP/6-31++G(3df,3pd) energies calculated at the CPCM-B3LYP/6-31++G(3df,3pd) optimized structures and including the zero-point energy, thermal and entropy terms from the B3LYP/6-31G* harmonic frequencies. The raw pK$_a$'s derived from these predicted free energies are listed in the top row of Tables 2-5 and plotted against the experimental pK$_a$'s in Figure 2. Following equation 3 above, note that the experimentally derived pK$_a$ for acetic acid (pK$_a$=4.76) was used as the reference compound for phosphoric acid and the four carboxylic acids. Neutral guanine (pK$_a$=3.3) and imidazole (pK$_a$=7.05) were used as the reference compounds for guanine and the four imidazoles. This plot also shows the least squares regression of all predicted pK$_a$'s taken together to yield a correlation coefficient (R^2) of 0.92, indicating a reasonably strong linear relationship. Taken separately, the linear fits most classes of compounds are much stronger, yielding . ....s of 0.98, 1.00, 0.83, and 0.98 for guanine, phosphates, imidazoles, and carboxylic acids, respectively.

Despite these strong linear fits, the predicted pK$_a$ values have a systematic shift in their range and magnitude. The experimental pK$_a$ values span a range of 13.3, 10.1, 3.1, and 4.5 for guanine, phosphoric acid, imidazoles, and carboxylic acids, respectively. The pKa values determined from our most computationally intensive calculation give somewhat greater ranges of 23.9, 19.7, 5.3, and 8.6. These systematic differences are evident in the best fits linear equations given in the top row of Table 6, mapping the predicted pK$_a$ values to the experimental values. The slopes, which should be 1

if our prediction was exact, instead range from 0.47 to 0.54. The nature of this systematic error in the slope is understandable because of the inherent errors in comparing neutral and ionic energies calculated using finite basis set quantum chemical methods. This is most clear for the calculations of successive pK_a's for compounds with several protonatable sites. For example, the phosphates range from the neutral phosphoric acid (H_3PO_4) to the trianionic phosphate (PO_4^{3-}). The more diffuse electronic wave function of the polyanions will be less accurately described by the basis set optimized for much less diffuse neutral atoms. Therefore, there will be a systematic underestimation in the energies of the more negatively charged compounds, which will have the effect of a bias towards the much larger protonation energies and therefore pK_a's. This bias is much greater in the gas-phase protonation energies (see discussion on *gas-phase energies* below).

Effect of neglecting enthalpy corrections and entropy

Comparing the pKa results derived from the predicted free energies, to those excluding ZPE, thermal corrections and entropy, we find that the use of the aqueous-phase electronic energies alone did not result in a significant decrease in accuracy. The predicted pK_a values derived from the electronic energies calculated with the CPCM-B3LYP/6-31++G(3df,3pd) at the CPCM-B3LYP/6-31++G(3df,3pd) structures are listed in the second row of Tables 2-5. A comprehensive plot of predicted versus experimental pK_a values for all compounds studied produced a slightly better linear fit with R^2 equal 0.94. An improvement in linear fit was also observed in carboxylic acids (R^2 = 1.00) whereas phosphates resulted in a very minor decrease in R^2 to 0.99, and g   no change in R^2. Comparing the slopes of the linear regression equations in the first and second rows of Table 6, pK_a values predicted from electronic energies were almost identical to those derived from predicted free energies. No difference was seen in the slopes for guanine and phosphoric acid, but a slightly smaller slope was seen in carboxylic acids (from 0.54 to 0.53). Comparing the range and magnitude of raw predicted pK_a values, there was also very little change between the two methods: no difference in range for guanine and phosphoric acid, but a difference of 0.1 for carboxylic acids. The negligible effect on the accuracy of the pKa values of including ZPE, thermal corrections and

115

entropy leads us to conclude that the time-consuming harmonic frequency calculation required to calculate these terms is unnecessary for such predictions.

Effect of using gas-phase structures

The calculation of electronic energies using aqueous-phase energies with gas-phase structures instead of aqueous-phase energies and structures resulted in an only a very small decrease in accuracy. The predicted pK_a values derived from the electronic energies calculated with CPCM-B3LYP/6-31++G(3df,3pd) at the gas-phase B3LYP/6-31++G(3df,3pd) structures are listed in the third row of Tables 2-5. There was no change in R^2 (0.98, 0.99, and 1.00, respectively) for guanine, phosphoric acid, or carboxylic acids. Comparing the slopes in the linear regression equations in the second and third rows of Table 6 showed an increase from 0.53 to 0.51 for both guanine and carboxylic acids and a decrease from 0.52 to 0.54 for phosphoric acid. Comparing the change in range and magnitude of predicted pK_a values, the differences ranged from as great as 0.9 to as small as 0.3 pKa units. The minor decrease in accuracy when using aqueous-phase energies with gas-phase structures indicates that at least for the compounds considered in this study, the solvent-induced change in molecular structure has an insignificant effect on the protonation energies. Note that there are examples where inclusion of aqueous solvation effects in the optimization are essential to get even qualitatively correct structures. For example, glycine (and the other amino acids) forms a zwitterion in aqueous solution, but in gas-phase quantum chemical optimizations using a sufficiently large basis set, the zwitterion will spontaneously rearrange to form the neutral.[2]

Effect of using a smaller basis set

The computational complexity of the DFT method used in the study scales between the third and fourth power of the size of the basis set. Hence, the choice of basis set greatly affects the computational time required. Considering for example the neutral guanine, a single B3LYP energy calculation required more than sixty times longer with the larger 6-31++G(3df,3pd) basis (508 functions) compared to the smaller 6-31G** basis (190 functions). Comparing the calculation of

pKa values from electronic energies using these two basis sets, we find that the smaller basis set actually yields a slightly better linear fit between the experimental and predicted pKa values. The pK_a values derived from the electronic energies calculated with CPCM-B3LYP/6-31* at the B3LYP/6-31G** structures can be found in the fifth row of Tables 2 and 5; only the carboxylic acids and guanine were not used to evaluate this method. While carboxylic acids resulted in no change in R^2, guanine experienced a slight increase from 0.98 to 0.99. In contrast to the previously discussed methods, comparing the slopes in the linear regression equations in the third and fifth rows of Table 6 gives somewhat larger systematic shifts from 0.51 to 0.40 for guanine and 0.51 to 0.45 for carboxylic acids. Comparing the span/range of raw predicted pK_a values, the differences ranged from as great as 7.5 to as small as 1.2. The increase in accuracy when using the smaller basis set gas-phase structures leads us to believe that this less sophisticated method is an adequate method to accurately calculate pK_a. But, since greater systematic shifts are observed, they need to be accounted for in the correction factors.

Effect of using Langevin dipole solvation model

We evaluated Langevin dipole method, both iterative and non-iterative, to calculate the solvation energy for the carboxylic acids. We found that this alternative solvation model yielded pK_a values which were closer in absolute magnitude to the experimental values, but which yielded a poorer linear fit than the PCM methods, and therefore could not be as well systematically corrected. The predicted pK_a values derived from the electronic energies calculated with Langevin dipole and HF/6-31G* at the B3LYP/6-31G** structures can be found in the eighth row of Table 5. Both non-iterated and iterated LD give poor R^2 values of 0.84 and 0.72, respectively. But, comparing the slopes in the linear regression equations in the eighth row of Table 6 gives systematic shifts of 0.86 and 1.07, our closest values for an exact slope. Comparing the predicted and experimental pK_a values, the differences ranged from as great as 4.9 to as small as 3.5.

Effect of using gas-phase energies

Facile gain and loss of protons does not occur in the gas-phase because of the great energy cost to separate charges in a non-polar environment. Nevertheless, gas-phase energies could be used to calculate the aqueous phase pKa values provided that there is a constant proportionality between the gas- and aqueous-phase protonation energies. Note that there is no a priori reason to expect this to always be accurate, and there are some examples where the gas-phase and aqueous-phase basicities are qualitatively different, such as the methylamines.[23, 24] The predicted pK_a values derived from the gas-phase electronic energies calculated with the B3LYP/6-31++G(3df,3pd) at the B3LYP/6-31++G(3df,3pd) structures are listed in the fourth row of Tables 2-5. In comparing the linear fit of these gas-phase pKa values, we find an actual increase in quality of this fit from R^2 equal 0.98 to 1.00, and from 0.99 to 1.00 for guanine and phosphoric acid, respectively. No change was seen in carboxylic acids. The slopes of the linear regression equations in the fourth row of Table 6 showed considerable increases in systematic shifts: from 0.51 to 0.06 for guanine, 0.54 to 0.06 for phosphoric acid, and 0.51 to 0.26 for carboxylic acids. The great increase in systematic shifts is also supported by the great range in predicted pK_a values (224.3, 181.4, and 18.0) versus experimentally derived pK_a values (13.3, 10.1, and 4.5) for guanine, phosphoric acid, and carboxylic acids. Even though these major shifts were present, correlation coefficients of very close to 1 maintains that pK_a values derived from the electronic energies calculated with the B3LYP/6-31++G(3df,3pd) at the B3LYP/6-31++G(3df,3pd) structures is a reliable method. But since much greater systematic shifts are observed, they must be accounted for in the correction factors. Also, since the shifts are so great, this method is more sensitive to errors.

Comparison to earlier work

A number of earlier studies have addressed the prediction of pK_a's for some of the chemical types investigated here. Schuurman et al. studied the prediction of pK_a's for a set of 16 aliphatic carboxylic acids [25], two of which are included here (acetic acid and fluoroacetic acid). He used Hartree-Fock (HF), HF including entropic and thermochemical corrections, and second order Moller-Plesset perturbation theory (MP2) with 6-31G**, 6-31+G**, 6-311G(2d, 2p) and 6-311+G(2d, 2p)

basis sets. To determine solvation effects, Schuurman used a united atom Hartree-Fock parameterization of PCM solvation model[26] to calculate the aqueous-phase energies at the HF gas-phase optimizes geometries. Comprehensively comparing their methods, they found the most reliable methods to accurately calculate pK_a were aqueous-phase HF excluding enthalpy and entropy terms at the 6-31G* and 6-31+G** basis sets ($R^2=0.93$). Our results support Schuurman's conclusion that smaller basis set calculations predict pK_a slightly more accurately than larger basis sets. Their findings are consistent with our result that use of gas-phase optimized structures is sufficient to accurate pKa predictions.

The pK_a's for orthophosphate have been calculated by Colvin et al. [27] who used second order Moller-Plesset perturbation theory at the Hartree-Fock/6-311++G** optimized geometries combined with PCM solvation model [21] based closely on the original method of Tomasi [28]. They calculated the pK_a's of both ortho- (H_3PO_4) and pyrophosphate ($H_7P_2O_7$) from equation 1 above, using the experimental value for the solvation energy of the proton. Their method yielded a somewhat better slope than we found between the predicted and experimental pK_a's for phosphate, but had a larger shift in absolute magnitude of about 6 pK_a units. Interestingly, for the larger pyrophosphate, their slope and offset was much closer to the values we found for our four sets of compounds.

Topol et al. have published a detailed study on the calculation of the pK_a values for several substituted imidazoles [29]. He used a dielectric continuum model and a variety of quantum chemical methods including HF, B3LYP-DFT, as well as very highly electron correlated coupled cluster methods and quadratic configuration interaction methods. Their study included two compounds, imidazole and 2-amino-imidazole, described in this paper. Topol found very little difference (<1 pK_a unit) between the pK_a values calculated using the B3LYP method (with a 6-311+G(d,p) basis set) and the much more computationally intensive methods. However, they found very large errors in the pK_a values when using HF/6-31G* energies, suggesting the need for electron correlation in the calculations. Overall, their solvent model led to slightly more accurate pK_a values than we found. In

agreement with our results, they predicted the pK_a for 2-amino-imidazole to be several pK_a units greater than the experimental value.

We are not aware of any previous aqueous-phase calculations of the pKa values for guanine. There have been reported a number of gas-phase proton affinities and aqueous-phase studies of the pKa relative to other DNA basis. Chandra, et al. very recently published gas-phase B3LYP/6-31+G(d,f) protonation and deprotonation enthalpies for guanine and adenine [30] . These results found that for neutral guanine in the gas-phase, the N_7 position is the most basic and the N_9 position is the most acidic, in agreement with earlier gas-phase predictions by Russo et al [31] and Colominas et al. [32].

Conclusion

In a study of a number of *ab initio* quantum chemical methods, we find a strong linear relationship when plotting the predicted and calculated pK_a's across a set of compounds and an even stronger fit within each class of compounds. Since each compound class yields somewhat different systematic shifts, correction factors must be recalibrated for each class of compounds to achieve greatest accuracy. We studied the effect of excluding the enthalpy and entropy terms, using the gas-phase optimized structures, and using a smaller basis set, and found in all cases no significant decrease in the accuracy of the predicted pK_a values. We also calculated the pK_a values from the gas-phase energies, excluding any solvation effects, and also found a strong linear relation with the experimental pK_a values, although with much larger systematic errors in the range and magnitude. Finally, we found that the Langevin dipole method yielded pKa values with sm..... a. ...lute e ....s than the PCM methods, but yielded poorer linear fits to the experimental values.

Acknowledgements:

References:

[1]   L. Stryer, Biochemistry, W. H. Freeman and Company, New York 1988.

[2]   M. P. Gamcsik, S. M. Ludeman, E. M. Shulman-Roskes, I. J. McLennan, M. E. Colvin, O. M. Colvin, Journal of Medicinal Chemistry, 36 (1993) 3636.

[3]   J. Tomasi, M. Persico, Chemical Reviews, 94 (1994) 2027.

[4]   A. A. Rashin, B. Honig, Journal of Physical Chemistry, 89 (1985) 5588.

[5]   M. Bucher, T. L. Porter, Journal of Physical Chemistry, 90 (1986) 3406.

[6]   M. W. Wong, M. J. Frisch, K. B. Wiberg, Journal of the American Chemical Society, 113 (1991) 4776.

[7]   M. W. Wong, M. J. Frisch, K. B. Wiberg, Journal of the American Chemical Society, 114 (1991) 523.

[8]   R. R. Pappalardo, E. S. Marcos, M. F. Ruiz-Lopez, D. Rinaldi, J.-L. Rivail, Journal of the American Chemical Society, 115 (1993) 3722.

[9]   C. J. Cramer, D. G. Truhlar, Journal of Computer-Aided Molecular Design, 6 (1992) 629.

[10]   V. Luzhkov, A. Warshel, Journal of Computational Chemistry, 13 (1992) 199.

[11]   B. Honig, K. Sharp, A.-S. Yang, Journal of Physical Chemistry, 97 (1993) 1101.

[12]   C. J. Cramer, D. G. Truhlar, Chemical Reviews, 99 (1999) 2161.

[13]   S. W. Rick, B. J. Berne, Journal of the American Chemical Society, 116 (1994) 3949.

[14]   A. H. De Vries, P. T. Van Duijnen, A. H. Juffer, International Journal of Quantum Chemistry, Symposium, 27 (1993) 451.

[15]   J. Florian, A. Warshel, Journal of Physical Chemistry B, 101 (1997) 5583.

[16]   R. G. Pearson, Journal of the American Chemical Society, 108 (1986) 6109.

[17]   F. Jensen, Introduction to Computational Chemistry, John Wiley and Sons, New York 1999.

[18]   R. F. Hout Jr., B. A. Levi, W. J. Hehre, Journal of Computational Chemistry, 3 (1982) 234.

[19]   A. D. Becke, Journal of Chemical Physics, 98 (1993) 5648.

[20]   C. Lee, W. Yang, R. G. Parr, Physical Review B, 37 (1988) 785.

[21]   V. Barone, M. Cossi, Journal of Physical Chemistry A, 102 (1998) 1995.

[22]   Y. Ding, Krogh-Jespersen, Chemical Physics Letters, 199 (1992) 261.

[23] D. H. Aue, H. M. Webb, M. T. Bowers, Journal of the American Chemical Society, 98 (1976) 311.

[24] D. H. Aue, H. M. Webb, M. T. Bowers, Journal of the American Chemical Society, 98 (1976) 318.

[25] G. Schüürmann, M. Cossi, V. Barone, J. Tomasi, Journal of Physical Chemistry A, 102 (1998) 6706.

[26] V. Barone, M. Cossi, J. Tomasi, Journal of Chemical Physics, 107 (1997) 3210.

[27] M. E. Colvin, E. M. Evleth, Y. Akacem, Journal of the American Chemical Society, 117 (1995) 4357.

[28] S. Miertus, E. Scrocco, J. Tomasi, Chemical Physics, 55 (1981) 117.

[29] I. A. Topol, G. J. Tawa, S. K. Burt, A. A. Rashin, Journal of Physical Chemistry A, 101 (1997) 10075.

[30] A. K. Chandra, M. T. Nguyen, T. Uchimaru, T. Zeegers-Huyskens, Journal of Physical Chemistry, A, 103 (1993) 8853.

[31] N. Russo, M. Toscano, A. Grand, F. Jolibois, Journal of Computational Chemistry, 19 (1997) 989.

[32] C. Colominas, F. J. Luque, M. Orozco, Journal of the American Chemical Society, 118 (1996) 6811.

Captions for All Figures and Tables

Figure 1. The protonated forms of the 10 compounds used in this study and their relative pK_a.values **a)** The 4 carboxylic acids and their respective, experimentally derived, pK_a's. **b)** Phosphoric acid. Fully protonated phosphoric·acid has a pK_a of 2.12 whereas dianionic phosphate has a pK_a of 12.32. **c)** The 4 imidazoles and their respective, experimentally derived, pK_a's. **d)** Guanine. The most acidic site for guanine is at the 3 position, with a pK_a of -1. The most basic site for guanine is at the 9 position, with a pK_a of 12.3.

Table 1. In this study, we evaluated the accuracy of several methods to calculate pK_a's. In all cases, density functional theory with a B3LYP functional was used. We then varied four parameters: comparing the inclusion of enthalpy and entropy terms (ΔG) in the electronic energy, the use of a smaller 6-31G** basis set versus the much larger 6-31++G(3df,3pd) basis, and the comparison of solvation effects.

Figure 2. A plot of experimentally derived pK_a's versus predicted pK_a's using the most computationally costly method (ΔG) for calculating molecular energies evaluated in this study. This plot yields an R^2 of 0.92. When separate plots are performed for each compound class, the linear fits are much stronger, with R^2 values of 0.98, 1.00, 0.83, and 0.98 for guanine, phosphates, imidazoles, and carboxylic acids, respectively. The raw pK_a's derived from these predicted free energies are listed in the top row of Tables 2-5.

Table 2. Table shows raw predicted pK_a's for guanine using,, from top to bottom: free energies (G) derived from large basis set CPCM-B3LYP/6-31++G(3df,3pd) energies calculated at the CPCM-B3LYP/6-31++G(3df,3pd) optimized structures and including the zero-point energy, thermal and entropy terms from the B3LYP/6-31G* harmonic frequencies, same method excluding enthalpy and entropy terms, aqueous-phase energies with gas-phase structures, gas-phase optimizations, and aqueous-phase energies with gas-phase structures at the smaller basis set. Experimentally derived pK_a values for all guanines are listed in the bottom row. Neutral guanine in the 7 position (p   was used as the reference compound.

Table 3. Table shows raw predicted pK_a's for the four protonation states of phosphoric acid using,, from top to bottom: free energies (G) derived from large basis set CPCM-B3LYP/6-31++G(3df,3pd) energies calculated at the CPCM-B3LYP/6-31++G(3df,3pd) optimized structures and including the zero-point energy, thermal and entropy terms from the B3LYP/6-31G* harmonic frequencies, same method excluding enthalpy and entropy terms, aqueous-phase energies with gas-phase structures, and gas-phase optimizations. Experimentally derived pK_a values for all phosphates are listed in the bottom row. Acetic acid (pK_a=4.76) was used as the reference compound.

Table 4. Table shows raw predicted pK$_a$'s for all imidazoles using,, from top to bottom: free energies (G) derived from large basis set CPCM-B3LYP/6-31++G(3df,3pd) energies calculated at the CPCM-B3LYP/6-31++G(3df,3pd) optimized structures and including the zero-point energy, thermal and entropy terms from the B3LYP/6-31G* harmonic frequencies, same method excluding enthalpy and entropy terms, aqueous-phase energies with gas-phase structures, and gas-phase optimizations, Experimentally derived pK$_a$ values for all imidazoles are listed in the bottom row. Imidazole (pK$_a$=7.05) was used as the reference compound.

Table 5. Table shows raw predicted pK$_a$'s for all carboxylic acids using,, from top to bottom: free energies (G) derived from large basis set CPCM-B3LYP/6-31++G(3df,3pd) energies calculated at the CPCM-B3LYP/6-31++G(3df,3pd) optimized structures and including the zero-point energy, thermal and entropy terms from the B3LYP/6-31G* harmonic frequencies, same method excluding enthalpy and entropy terms, aqueous-phase energies with gas-phase structures, gas-phase optimizations, aqueous-phase energies with gas-phase structures at the smaller basis set, and Langevin dipole models. Experimentally derived pK$_a$ values for all carboxylic acids are listed in the bottom row. Acetic acid (pK$_a$=4.76) was used as the reference compound.

Table 6. Table shows the linear regression equations and correlation coefficients for all compounds and methods evaluated.

Figure 1.

Carboxylic acids

Acetic acid
4.76

Fluoroacetic acid
2.66

Difluoroacetic acid
1.24

Trifluoroacetic acid
0.23

(a)

Phosphoric acid

2.12, 7.21, 12.32

(b)

Imidazoles

Imidazole
7.05

2-aminoimidazole
8.46

Benzimidazole
5.4

2-aminobenzimidazole
7.51

(c)

Guanine

3.3

12.3

9.2

-1

(d)

Comprehensive Plot, ΔG

Calculated pKa

Experimental pKa

Legend:
- ◆ Carboxylates
- ◀ Guanines
- ● Phosphates
- ✻ Imidazoles
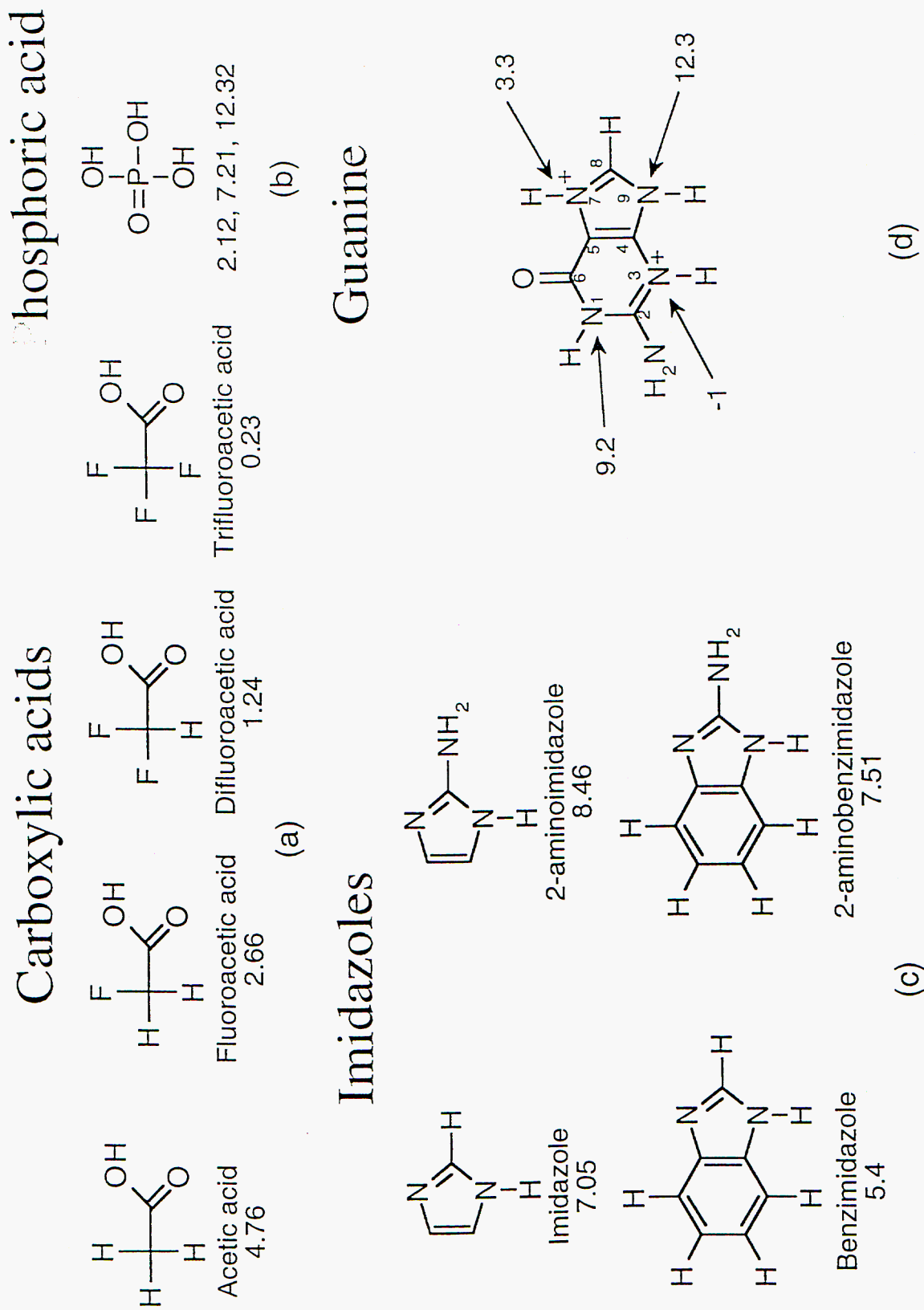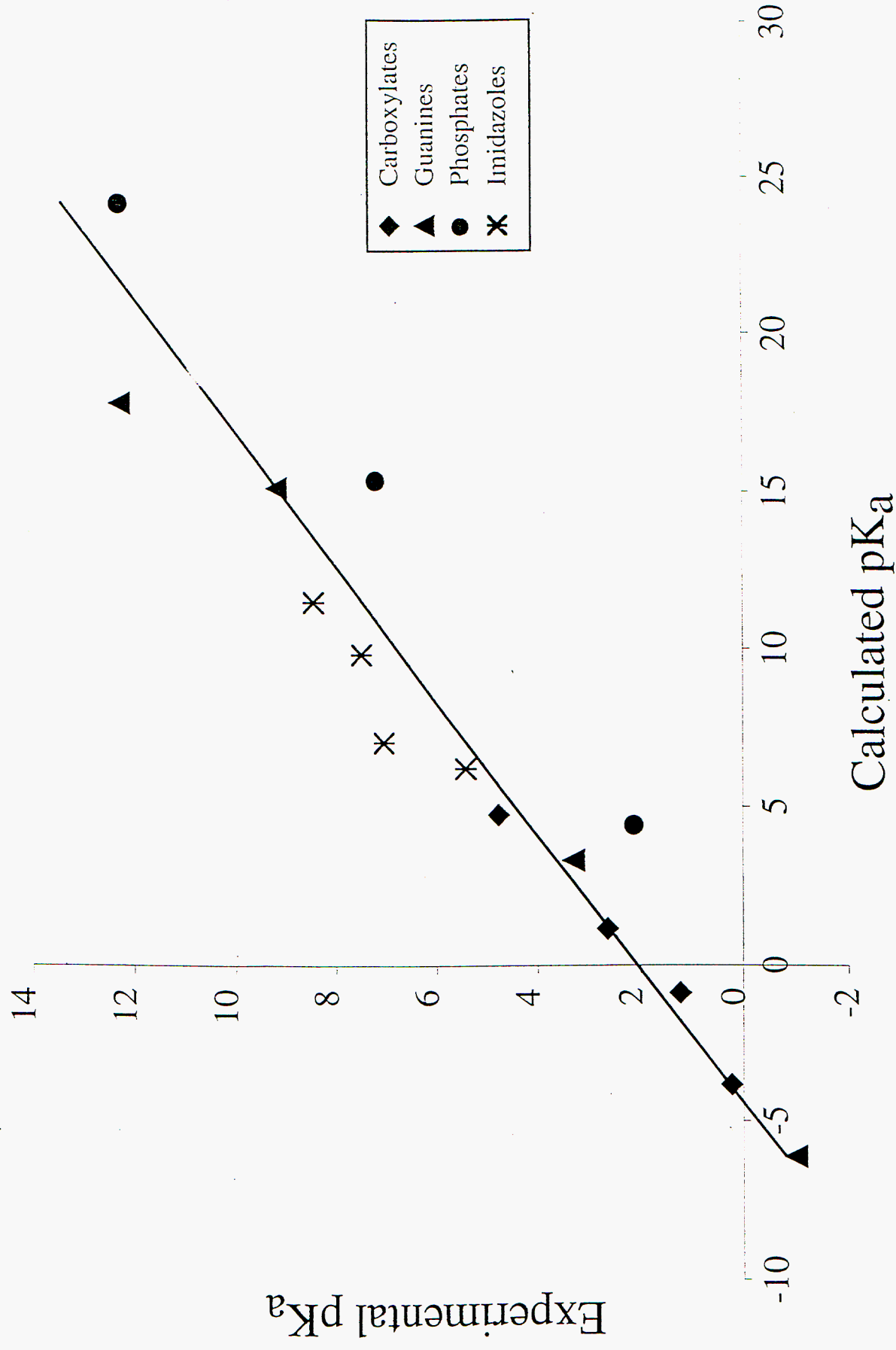
Table 1.

---	Options
Energy Terms	Electronic energy (ΔE) **or** Free energy (ΔG)
Basis Set	6-31G** **or** 6-31++G(3df,3pd)
Solvent Model	None (Gas-phase) **or** PCM **or** LD
Optimization	Gas-phase **or** Aqueous-phase

Table 2.

Guanines

Methods	9guanine	7guanine	3guanine	1guanine
CPCM-B3LYP/6-31++G(3df,3pd)// CPCM-B3LYP/6-31++G(3df,3pd) + ZPE - TΔS	17.81	15.14	*3.3*	-6.12
CPCM-B3LYP/6-31++G(3df,3pd)// CPCM-B3LYP/6-31++G(3df,3pd)	17.71	15.56	*3.3*	-6.15
CPCM-B3LYP/6-31++G(3df,3pd)// B3LYP/6-31++G(3df,3pd)	17.61	15.36	*3.3*	-7.07
B3LYP/6-31++G(3df,3pd)// B3LYP/6-31++G(3df,3pd)	147.39	83.9	*3.3*	-76.87
CPCM-B3LYP/6-31G**// B3LYP/6-31G**	23.71	19.28	*3.3*	-8.46
Experiment	*12.3*	*9.2*	*3.3*	*-1*

Table 3.

Phosphates

Methods	Monohydro phosphate	Dihydro phosphate	Phosphoric acid
CPCM-B3LYP/6-31++G(3df,3pd)// CPCM-B3LYP/6-31++G(3df,3pd) + ZPE - TΔS	24.18	15.37	4.45
CPCM-B3LYP/6-31++G(3df,3pd)// CPCM-B3LYP/6-31++G(3df,3pd)	22.56	14.02	2.88
CPCM-B3LYP/6-31++G(3df,3pd)// B3LYP/6-31++G(3df,3pd)	21.66	14.27	2.9
B3LYP/6-31++G(3df,3pd)// B3LYP/6-31++G(3df,3pd)	174.88	83.24	-6.56
Experiment	*12.32*	*7.21*	*2.2*

Table 4.

Imidazoles

Methods	imidazole	2amino-imidazole	benzimidazole	2amino-benzimidazole
CPCM-B3LYP/6-31++G(3df,3pd)// CPCM-B3LYP/6-31++G(3df,3pd) + ZPE - TΔS	*7.05*	11.52	6.19	9.86
CPCM-B3LYP/6-31++G(3df,3pd)// CPCM-B3LYP/6-31++G(3df,3pd)	*7.05*	10.38	6.14	8.77
CPCM-B3LYP/6-31++G(3df,3pd)// B3LYP/6-31++G(3df,3pd)	*7.05*	12.84	6.11	9.09
B3LYP/6-31++G(3df,3pd)// B3LYP/6-31++G(3df,3pd)	*7.05*	12.45	8.99	12.44
Experiment	*7.05*	*8.46*	*5.4*	*7.51*

Table 5.

Carboxylates

Methods	Acetic acid	Fluoro-acetic acid	Difluoro-acetic acid	Trifluoro-acetic acid
CPCM-B3LYP/6-31++G(3df,3pd)// CPCM-B3LYP/6-31++G(3df,3pd) + ZPE - TΔS	4.76	1.1	-.88	-3.84
CPCM-B3LYP/6-31++G(3df,3pd)// CPCM-B3LYP/6-31++G(3df,3pd)	4.76	-.58	-1.61	-3.94
CPCM-B3LYP/6-31++G(3df,3pd)// B3LYP/6-31++G(3df,3pd)	4.76	.44	-1.71	-4.24
B3LYP/6-31++G(3df,3pd)// B3LYP/6-31++G(3df,3pd)	4.76	-3.16	-8.32	-13.24
CPCM-B3LYP/6-31G**// B3LYP/6-31G**	4.76	.23	-2.72	-5.45
LD + non-iter//HF/6-31G*	4.76	3.6	2.71	-0.15
LD + solv//HF/6-31G*	4.76	4.27	3.74	1.25
Experiment	*4.76*	*2.66*	*1.24*	*.23*

Table 6.

Methods	Guanines	Phosphates	Imidazoles	Carboxylates
CPCM-B3LYP/6-31++G(3df,3pd)// CPCM-B3LYP/6-31++G(3df,3pd) + ZPE - TΔS	Y=0.53x+1.93 R^2=0.98	Y=0.52x-0.34 R^2=1.00	Y=0.47x+3.01 R^2=0.83	Y=0.54x+2.07 R^2=0.98
CPCM-B3LYP/6-31++G(3df,3pd)// CPCM-B3LYP/6-31++G(3df,3pd)	Y=0.53+1.93 R^2=0.98	Y=0.52x+0.44 R^2=0.99	Y=0.64x+1.94 R^2=0.88	Y=0.53x+2.25 R^2=1.00
CPCM-B3LYP/6-31++G(3df,3pd)// B3LYP/6-31++G(3df,3pd)	Y=0.51x+2.20 R^2=0.98	Y=0.54x+0.28 R^2=0.99	Y=0.39x+3.73 R^2=0.81	Y=0.51x+2.32 R^2=1.00
B3LYP/6-31++G(3df,3pd)// B3LYP/6-31++G(3df,3pd)	Y=0.06x+3.55 R^2=1.00	Y=0.06x+2.50 R^2=1.00	Y=0.29x+4.15 R^2=0.36	Y=0.26x+3.49 R^2=1.00
CPCM-B3LYP/6-31G**// B3LYP/6-31G**	Y=0.40x+2.17 R^2=0.99	-----	-----	Y=0.45x+2.58 R^2=1.00
LD + non-iter//HF/6-31G*	-----	-----	-----	Y=0.86x-0.12 R^2=0.84
LD + solv//HF/6-31G*	-----	-----	-----	Y=1.07x-1.51 R^2=0.72